

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 12:44:51 ; Search time 820 Seconds
(without alignments)
10672.302 Million cell updates/sec

Title: US-09-831-804-1

Perfect score: 2060
Sequence: 1 cttattagggaagatgct.....gacattactctctaatg 2060

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_genseq29Jan04:*
2: genseq1980s:*
3: genseq1990s:*
4: genseq2000s:*
5: genseq2001as:*
6: genseq2002s:*
7: genseq2003as:*
8: genseq2003bs:*
9: genseq2003cs:*
10: genseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2060	100.0	2060	3	AA15398 DNA encod
2	1235.8	60.0	1239	6	AB232206 Candida a
3	99.4	4.8	2132	7	ADA53158 Human cod
4	98.4	4.8	1705	8	ACA98970 CDNA enco
5	96.6	4.7	2320	7	ACA56456 Human sig
6	95.2	4.6	4563	5	AA591317 DNA encod
7	94.6	4.6	2597	7	AA151569 Human nuc
8	92.4	4.5	3639	5	AA664586 DNA encod
9	91.8	4.5	2110	9	ADB63624 Human cod
10	91.8	4.5	2760	9	ADC56695 Human mac
11	91.8	4.5	3078	9	ADB62468 Human cod
12	90.4	4.4	2026	7	ADB5863 Human nuc
13	90.4	4.4	2114	7	ADA53124 Human cod
14	90.2	4.4	2230	7	ADA52931 Human cod
15	90	4.4	1952	9	ADC58104 Zinc fing
16	89.2	4.3	2064	9	ADC30762 Human nov
17	89.2	4.3	2622	5	AA68872 DNA encod
18	89.2	4.3	2729	6	AA16178 Human cod
19	89.2	4.3	3839	6	AAK83826 Human cod
20	89	4.3	2298	4	AA16608 Human cod
21	89	4.3	2905	4	AA56143 DNA encod
22	89	4.3	3020	5	AA592560 DNA encod
23	89	4.3	3502	4	AA157845 Human pol

24	88.6	4.3	2597	7	AAU55855	AA55855 Human nuc
25	88.6	4.3	4227	8	ACA98938	ACA98938 cDNA enc
26	88.4	4.3	2509	6	AAU31103	AA311103 Human tra
27	87.6	4.3	1757	6	ABO81135	ABq81135 TRAF6-Inh
28	87.2	4.2	831	6	ABQ5095	ABq5095 Human ova
29	87	4.2	976	5	AA69188	AA69188 DNA encod
30	86.8	4.2	2662	6	ABQ93353	ABq93353 Human cod
31	86.8	4.2	6219	7	ACC46324	ACC46324 Human dit
32	86.8	4.2	6316	7	ABX34443	ABx34443 Human mdd
33	86.2	4.2	1777	7	AA62690	AA62690 Human gen
34	86.2	4.2	2558	4	AA526691	AA526691 Human gen
35	86.2	4.2	2558	4	ABX74039	ABx74039 Human nov
36	86.2	4.2	2558	7	AA587125	AA587125 DNA encod
37	86.2	4.2	2607	5	ABX34772	ABx34772 Human mdd
38	86.2	4.2	2681	7	ACC40336	ACC40336 Human nov
39	86	4.2	1890	9	ACC46347	ACC46347 Human dit
40	86	4.2	3309	7	ADB62883	ADB62883 Human cod
41	86	4.2	3400	9	AA667562	AA667562 DNA encod
42	86	4.2	1549	5	ACC46407	ACC46407 Human dit
43	85.6	4.2	2476	7	ADA53516	ADa53516 Human cod
44	85.6	4.2	2476	7		
45	84.8	4.1	2239	7		

ALIGNMENTS

RESULT 1

AA15398
ID AA15398 standard; DNA; 2060 BP.

AC AA15398;

DT 04-SEP-2000 (first entry)

DE DNA encoding a transcription factor designated CATFIIIA.

KW Transcription factor; CATFIIIA; DNA-binding protein;

XW ribosomal RNA 5S gene; fungal infection; ss.

XX

OS Candida albicans.

XX Key Location/Qualifiers

XX CDS 720..1958

XX 18-MAY-2000.

XX 09-NOV-1999; 99WO-FR002739.

XX 10-NOV-1998; 98FR-00014147.

XX (HMRI) HOECHST MARION ROUSSEL.

XX Bordon-Pallier F, Camier S, Sentenac A;

XX WPI; 2000-376549/32.

XX P-PSDB; AAY93316.

XX New nucleic acid encoding Candida albicans transcription factor, useful

XX e.g. in screening for antimycotic agents and for immunization.

XX Claim 4; Page 32-33; 45pp; French.

XX The present sequence encodes a Candida albicans transcription factor, designated CATFIIIA. The polypeptide is a DNA-binding protein, which is

involved in initiating transcription of the ribosomal RNA 5S gene. The polynucleotide is used to screen for its specific inhibitors, potentially

useful as antimycotic agents, to raise an antibody response that is

CC protective against fungal infection to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in
CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations

XX Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 2060; DB 3; Length 2060;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2060; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTTTATAGGAATGGCTGAGGCCATTTTGTATTAACGGGCTCCCAAGTGAATGTTT 60
DB 1 CTTTATAGGAATGGCTGAGGCCATTTTGTATTAACGGGCTCCCAAGTGAATGTTT 60
QY 61 TAGTAATATCCATCATTTGGGCTTCAAGTGAATGGGGTGTCTCATCTCTTGTGTAG 120
DB 61 TAGTAATATCCATCATTTGGGCTTCAAGTGAATGGGGTGTCTCATCTCTTGTGTAG 120
QY 121 AAATAGGCGCAGGCTCCGAAATCCAAAAGAAAGAAATCAGATGTCCTGGCTGCAAG 180
DB 121 AAATAGGCGCAGGCTCCGAAATCCAAAAGAAAGAAATCAGATGTCCTGGCTGCAAG 180
QY 181 TTTGTGCGCATGGCAATGCCGAAATGAAAAAAGTCTACTGGGCCCACT 240
DB 181 TTTGTGCGCATGGCAATGCCGAAATGAAAAAAGTCTACTGGGCCCACT 240
QY 241 ACAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 ACAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 TATTTGCAAGGCTTTAGAAATTTGTATATTTGTGTCTGACACTGTGTGTATATCT 360
DB 301 TATTTGCAAGGCTTTAGAAATTTGTATATTTGTGTCTGACACTGTGTGTATATCT 360
QY 361 GGCATCTGCTGCTCCCTGGAAGGGCTGCTGAATGAATCATGATCAAGATATAT 420
DB 361 GGCATCTGCTGCTCCCTGGAAGGGCTGCTGAATGAATCATGATCAAGATATAT 420
QY 421 GACTTGTCTCACTTATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 421 GACTTGTCTCACTTATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 481 TATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 481 TATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 CAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 541 CAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 601 ATTTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 601 ATTTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 AAAAATTAACGCGGACTCATCTCTTTTCAAAATCCATTTTATATCTTCTTCAATCA 720
DB 661 AAAAATTAACGCGGACTCATCTCTTTTCAAAATCCATTTTATATCTTCTTCAATCA 720
QY 721 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 721 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 GTCCCAAAAAGTATATTTGACATATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 781 GTCCCAAAAAGTATATTTGACATATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 TATTAAGCAACATTTAAGAACCAAGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 841 TATTAAGCAACATTTAAGAACCAAGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 ATTGTGATTAAGCATTTTTCGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 901 ATTGTGATTAAGCATTTTTCGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 960
```

```
DB 901 ATTGTGATTAAGCATTTTTCGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 AAAAAAACCATTTCCATTTGTCAGGTGTGTAAAGGGTTATTTCTGACAAACATTTGA 1020
DB 961 AAAAAAACCATTTCCATTTGTCAGGTGTGTAAAGGGTTATTTCTGACAAACATTTGA 1020
QY 1021 AAAGCATGAAATCCGCCATACAAAGTCAATTAATGATGATGATGATGATGATGATGATGAT 1080
DB 1021 AAAGCATGAAATCCGCCATACAAAGTCAATTAATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 CATTTTATTAACATCATCTTTTAAGCATCATATATATCTGTCTGATGAAAAACATTA 1140
DB 1081 CATTTTATTAACATCATCTTTTAAGCATCATATATATCTGTCTGATGAAAAACATTA 1140
QY 1141 CGGTAAACAAATGATTAAGTTTTCATGACCTTCAAAATTAAGCAACATTAATTA 1200
DB 1141 CGGTAAACAAATGATTAAGTTTTCATGACCTTCAAAATTAAGCAACATTAATTA 1200
QY 1201 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 AAACCTGTGATGATTAACATTTTCAATTAACAACTGATCCAAACCTTAATGTCCTA 1320
DB 1261 AAACCTGTGATGATTAACATTTTCAATTAACAACTGATCCAAACCTTAATGTCCTA 1320
QY 1321 AATGTGTAAAGTTGTGTGGAAAAAGTTATCTCAATGATGATGATGATGATGATGATGATGAT 1380
DB 1321 AATGTGTAAAGTTGTGTGGAAAAAGTTATCTTCAATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 ATTCATCATGATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1381 ATTCATCATGATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 AAAATGAATTTGTTGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
DB 1441 AAAATGAATTTGTTGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
QY 1501 TAAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
DB 1501 TAAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1561 ATTTGATGATTAAGAAACAGAGAAATTAAGTGAAGAAATGAAGAAATGAAGAAATGAAG 1620
DB 1561 ATTTGATGATTAAGAAACAGAGAAATTAAGTGAAGAAATGAAGAAATGAAGAAATGAAG 1620
QY 1621 ATAGCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
DB 1621 ATAGCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 TAAATCATTTACTGCTCTTGAAGGTTCAAGAGTGTCTTAACTTATCTGAATA 1740
DB 1681 TAAATCATTTACTGCTCTTGAAGGTTCAAGAGTGTCTTAACTTATCTGAATA 1740
QY 1741 GTGGGAAAGATCAATTTGCTTAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1800
DB 1741 GTGGGAAAGATCAATTTGCTTAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1800
QY 1801 ATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
DB 1801 ATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
QY 1861 ATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
DB 1861 ATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
QY 1921 ATTTATGCAAAATGAACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
DB 1921 ATTTATGCAAAATGAACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
QY 1981 CATTTTATTTCTTTAATTTTATTTTGTGGCTTTTATTTTATTTTATTTTATTTTATTTT 2040
DB 1981 CATTTTATTTCTTTAATTTTATTTTGTGGCTTTTATTTTATTTTATTTTATTTTATTTT 2040
```

QY 2041 GACATATCTCTTATG 2060
 Db 2041 GACATATCTCTTATG 2060

RESULT 2
 ABZ32206
 ID ABZ32206 standard; DNA; 1239 BP.

AC ABZ32206;
 DT 30-JAN-2003 (first entry)
 DE Candida albicans essential gene SEQ ID NO 6493.

XX Fungus; Yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
 KM signal transduction; DNA replication; cell division; growth;
 XX proliferation; Candida albicans; fungicide; antifungal; gene; ss.

OS Candida albicans.
 PN WO200253728-A2.

PD 11-JUL-2002.
 PF 26-DEC-2001; 2001WO-US049486.

PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00792024.
 PR 22-AUG-2001; 2001US-0314050P.

XX (ELIT-) ELITRA PHARM INC.
 PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
 DR MPI: 2002-566694/60.
 XX P-PSDB; ABP73656.

PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX
 PS Claim 37; SEQ ID NO 6493; 167bp + Sequence Listing; English.

XX The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX
 SQ Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 60.0%; Score 1235.8; DB 6; Length 1239;
 Best Local Similarity 99.8%; Pred. No. 9.7e-217;
 Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	720	ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTATATCTTCTCTTCATCA	779
Db	1	ATGAGTGAAGTGAAGCAAAACCAATGATATCATCTTTATATCTTCTCTTCATCA	60
QY	780	CGTCCCAAAAAGATATTTTGACATATGAAAGGTGATTAAGCTTAATGACATCA	839
Db	61	CGTCCCAAAAAGATATTTTGACATATGAAAGGTGATTAAGCTTAATGACATCA	120
QY	840	TTATTAGACCAATTTAAGAACCCAGTAATGATCGACCGTATTAAGTAAAGTAC	899
Db	121	TTATTAGACCAATTTAAGAACCCAGTAATGATCGACCGTATTAAGTAAAGTAC	180
QY	900	GATTGTGATTAAGCAATTTTTCAGAAATGCAATTTGAAACATATTTGATCATTCC	959
Db	181	GATTGTGATTAAGCAATTTTTCAGAAATGCAATTTGAAACATATTTGATCATTCC	240
QY	960	GAATAAAAAACCAATTCATTTGATGATGATGATTAAGGGGTATATTTCTGACAC	1019
Db	241	GAATAAAAAACCAATTCATTTGATGATGATGATTAAGGGGTATATTTCTGACAC	300
QY	1020	AAAAGCATGAATCAACCATCAAAAGTCAATTTAATGTACATTTGAAATTTGCAAGA	1079
Db	301	AAAAGCATGAATCAACCATCAAAAGTCAATTTAATGTACATTTGAAATTTGCAAGA	360
QY	1080	GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTA	1139
Db	361	GCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAACATTA	420
QY	1140	ACGTGTAACATGTATTAATTAAGTTTCACTGCACCTTCAAAATTTGACACATTAATTA	1199
Db	421	ACGTGTAACATGTATTAATTAAGTTTCACTGCACCTTCAAAATTTGACACATTAATTA	480
QY	1200	AAACATCATGTGATCTCTCTCTATCATATGATGATCATCTGTTTAAATTTTC	1259
Db	481	AAACATCATGTGATCTCTCTCTATCATATGATGATCATCTGTTTAAATTTTC	540
QY	1260	CAACTTGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1319
Db	541	CAACTTGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	600
QY	1320	AAATGTGTAAGGTTGTTGTTGGGAAAAAGTTTATCTTCAATATGTTAATGATGAT	1379
Db	601	AAATGTGTAAGGTTGTTGTTGGGAAAAAGTTTATCTTCAATATGTTAATGATGAT	660
QY	1380	GATTCTACATGATCAAAATATGACCTGTGATTAATGATGATGATGATGATGATGAT	1439
Db	661	GATTCTACATGATCAAAATATGACCTGTGATTAATGATGATGATGATGATGATGAT	720
QY	1440	AAAAATGATTAAGTTGAACATTATTAATATCTTCCATGATGATGATGATGATGAT	1499
Db	721	AAAAATGATTAAGTTGAACATTATTAATATCTTCCATGATGATGATGATGATGAT	780
QY	1500	TTAAAGGAACGTAAGTAAAAATTTGAGAACCTTAATTAATGATCAAGATGCAATTAAT	1559
Db	781	TTAAAGGAACGTAAGTAAAAATTTGAGAACCTTAATTAATGATCAAGATGCAATTAAT	840
QY	1560	AATTTGATGATTAAGGAACGAGAAATTTAAGGTGAAAGATGAAGAGATGAAGA	1619
Db	841	AATTTGATGATTAAGGAACGAGAAATTTAAGGTGAAAGATGAAGAGATGAAGA	900
QY	1620	GATGCTGATGATTAAGGAACGAGATGATGATGATGATGATGATGATGATGATGAT	1679
Db	901	GATGCTGATGATTAAGGAACGAGATGATGATGATGATGATGATGATGATGATGAT	960
QY	1680	ATTAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1739
Db	961	ATTAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1020
QY	1740	AGTGGGAAGAGATCAATTTGCTTAAGAAATTAATGATGATGATGATGATGATGAT	1799

```

Db      1021 AGTGGAAAGAAATCATATGTCCTTAAGATTAATGTGATAGAAATGTTTCTGAGAAATAT 1080
QY      1800 GATTACGTCGACATTTGAATGCGATGATGATTAATTTTCAAGAATTTGAGTCATTTCTTA 1859
Db      1081 GATTTACGTCGACATTTGAATGCGATGATGATTAATTTTCAAGAATTTGAGTCATTTCTTA 1140
QY      1860 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAAACCATTTGGTTAAAAAAGCCAGATG 1919
Db      1141 AATAGTATGAAAAAGAAAGAACTCCAGAAAGTGAAACCATTTGGTTAAAAAAGCCAGATG 1200
QY      1920 GATTTATTCGCAAAATGAAACATCAGTGAATTTCTCGATAA 1958
Db      1201 GATTTATTCGCAAAATGAAACATCAGTGAATTTCTCGATAA 1239

```

RESULT 3

ADAS3158
ID ADAS3158 standard; cDNA; 2132 BP.

AC ADAS3158;

XX 20-NOV-2003 (first entry)

DE Human coding sequence, SEQ ID 726.

KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW Inflammatory disease; osteoporosis; neurological disease; gene; ss.

OS Homo sapiens.

PN EP1293569-A2.

XX 19-MAR-2003.

PF 21-MAR-2002; 2002EP-00006586.

PR 14-SEP-2001; 2001JP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;

DR WPI: 2003-395539/38.

XX P-PSDB; ADA54797.

PT New polynucleotides encoding full-length polypeptides, e.g. secretory

PT and/or membrane proteins, useful for developing medicines for diseases in

XX which the gene is involved, or as target molecules for gene therapy.

PS Claim 1; SEQ ID NO 726; 205bp; English.

XX The present invention relates to novel human secretory or membrane

CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-

CC ADA54071). The coding sequences are useful in the gene therapy of

CC diseases caused by abnormalities of the proteins, e.g. cancer,

XX inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

Query Match 4.8%; Score 99.4; DB 7; Length 2132;

Best Local Similarity 50.6%; Pred. No. 6.6e-09;

Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 726 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTTCTTCTGATCAGTCGCC 785

Db 1384 GAATGTGGCAAAAGCTTTTAAACGAGTCCATCTTACTACTACATAAGAGATTCTACT 1443

```

QY      786 AAAAGTATATTTGACATATGAGAGGTGTGATTAAGACCTATATATGACCATCATTTATTA 845
Db      1444 GGAGAGAAATTCCTACAAATGTGAAGATGTGGCAAGCTTCTATGATCCCTCAAACTT 1503
QY      846 GAGCAATTTTAAAGAACCCAGATATGATGACCGGTATTAATGTAACAGTGAACGATTTGT 905
Db      1504 ACTGAACATPAGAAATTCATCTACTGAGAGAAACCCCTACACATGT-----GAGAGATGT 1557
QY      906 GATTAAGCATTTTTCAGAAATATCACATTTGGAACACATATGTATCACATTCGAAAAA 965
Db      1558 GGCAAAAGCCTTTTAAACCATCTCTACACCTTGCTACACATAGGTATTTCTACTCTGAGAG 1617
QY      966 AAACCATTCATTTGTTCAGTGTGTGTAAAGGGGTATTTCTGACACACCTTGAAGA 1025
Db      1618 AAACCTTACCATTTGTAAGAAATGTGTGTAAAGCCTTTAACAGATCTCACACCTTACTAG 1677
QY      1026 CATGAATACACCCATTCGAAAGTCAATTTAAATGTACATTTGAAATTTGTCAAGAACATTT 1085
Db      1678 CATAGAGAAATTCATPACTGAGAGAAACCCCTACCAATGTGAAAAAATGTGCAAAAGCTTTT 1737
QY      1086 TATTAACATCAATCTTTAA--GACATCAATATTTATCTGTTGATGAAAAAAGATTAAAG 1142
Db      1738 AACAGTCTCTCAAACTTACTGTGACATTAAGAAATTCATCTGTGAGAAACCTCAAAA 1797
QY      1143 TGTAAACAATGTATTAAGTTTCACTGACCTTCAAAATTTAGCACATTAATTTAAA 1202
Db      1798 CTTAAAAAGATGTAAACGTGATTTGAAAAACACTTCAAAAGTTTCTAAACATTAAGAAAT 1857
QY      1203 CATCATGTGTGATCTCTCTGTTTCAATGTGATCAATCTGTGTTTAAATTTCCAA 1262
Db      1858 TATGCTGTGTGAGAAATCTTAAAGAAATGTGAAGATGTGAACAAACCTTTAAAGTTGTAC 1917
QY      1263 ACTTGTGAGTATTTACATTTTCATATAAAACAATGCATCCAAAGT 1309
Db      1918 ACTTGAATGTGCTAATAGATTAATTCATCTAATTAAGT 1964

```

RESULT 4

ACA98970
ID ACA98970 standard; cDNA; 1705 BP.

AC ACA98970;

XX 25-JUL-2003 (first entry)

DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.

XX Human; nucleic acid-associated protein; cytosolic; anti-arteriosclerotic;

KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;

KW anti-allergic; anti-inflammatory; chymotrypsin; gene therapy;

KW cell proliferative disorder; cancer; atherosclerosis;

KW neurological disorder; epilepsy; Huntington's disease; stroke;

KW immune disorder; inflammatory disorder; AIDS; allergy;

KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;

KW protein-protein interaction; drug-target interaction;

XX gene expression profile; gene; ss.

XX Homo sapiens.

XX W02003023003-A2.

XX 20-MAR-2003.

XX 05-SEP-2002; 2002WO-US028540.

XX 07-SEP-2001; 2001US-0317792P.

XX 14-SEP-2001; 2001US-0317912P.

XX 21-SEP-2001; 2001US-0322270P.

XX 28-SEP-2001; 2001US-0324040P.

XX 21-SEP-2001; 2001US-0326732P.

XX 19-OCT-2001; 2001US-0346716P.

XX 25-JAN-2002; 2002US-0351749P.

XX 22-FEB-2002; 2002US-0359498P.

XX (INCY-) INCYTE GENOMICS INC.
 PA Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe J;
 XX Becha SD, Richardson TW, Lee EA, Sprague MW, Emerging BM;
 PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JK;
 PI Hafalia AJA, Santanwala B, Margulis JP, Gorvad AE, Lee SY, Ison CH;
 PI Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebatjadian Y, Shah P;
 PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kahle AE;
 PI Burford N, Ramkumar J;
 XX MPI: 2003-313243/30.
 DR P-PSDB; AB096722.
 XX
 PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
 PT treating and preventing diseases or conditions associated with the
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
 XX infections.
 PS Claim 5; Page 340-341; 345pp; English.
 XX
 CC The invention describes a novel human isolated nucleic acid-associated
 CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
 CC diagnosing, treating and preventing diseases or conditions associated
 CC with the decreased expression or overexpression of NAAP, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
 CC disorders, or infections. These are also useful in assessing the effects
 CC of exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of NAAP. The NAAP or its fragments are useful in screening
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,
 CC or in altering the expression of the target polynucleotide and compounds
 CC that specifically bind to or modulate the activity of the polypeptide.
 CC The microarray is useful in monitoring or measuring protein-protein
 CC interactions, drug-target interactions, and gene expression profiles.
 CC This sequence encodes a novel human nucleic acid-associated protein
 CC (NAAP)
 XX
 SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;
 Query Match 4.8%; Score 98.4; DB 8; Length 1705;
 Best Local Similarity 52.6%; Pred. No. 9.8e-09;
 Matches 266; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

710 ACTTATTCATGAGTGAAGTGAAGCAACCAATCGATATCATCTTATATCTTCTTC 769
 DB 679 AATTTTAAATGTAAAGATGTGCAAAAATTTTGCATGCTTTCACACTTAACACACA 738
 QY 770 TTCTTCATCAGCTCCCAAAAAGTATATTTGGCAGATGAGGGGTGATAAAGCCTATAA 829
 DB 739 TAAAGAAATCCAAAGCTAGAGTAAATTTTACAAATGTGAGCATATGAGAGCCTTAA 798
 QY 830 TCGACCATCATATTAAGAGCAATTTAAGAACCAAGCATATGATCGACCTATTAATG 889
 DB 799 CTGGTCCCAACCTTAATTAACATTAAGAAATTAATCTGAGAGAAAACCTTAACAAATG 858
 QY 890 TAGAGTGAAGATGTGATTAAGCAATTTTTCAGAAAATCACATTTGGAACACATATTTGT 949
 DB 859 TAAA-----GAATGTGGCAAGCCTTAACCAAGACCTCACCTTATTAAGCATTAAGAG 912
 QY 950 ATCAGATTCGGAATAAAACCATTCATTTTCAGTGTGTGTAAGGGGTTATTTCTCG 1009
 DB 913 AATTATATCTGAAGAAACCTTCAAAATGTGAAATGTGGCAAGCCTTAAACAGATC 972
 QY 1010 ACAACATTTGAAAAGACATGAATCACCATTAACCAAGTCAATTAATGATTAATGAAA 1069
 DB 973 ATCGACCTTACTACACATTAATTAATTAATCTACCTGGGAAATTCCTCAAAATGTGA 1032
 QY 1070 TTGTCAAGAGCATTTTAAACATCAATCTTTAAGACATCATATATATCTGTTCACGA 1129
 DB 1033 ATGTGTAGAGCTTTTAAACCAAGCTCAAGAGCTTACGTGAACATTAAGTTAATTCATACCGG 1092

QY 1130 A---AAACATTACGCTGTAAACATGATATATAAGTTTTCACCTGACCTTCNAATTAAGC 1186
 DB 1093 AGAGAAATGCTCAAAATGTGAAGATGTGGCAAAAGTTTAAACGATCTTCAAACTTAC 1152
 QY 1187 ACAACATTAATTTAAACATCAATGCTG 1212
 DB 1153 AGACCATTAATAAAATTCATACAGAG 1178

RESULT 5
 ACAS6456
 ID ACAS6456 standard; cDNA; 2320 BP.
 XX
 AC ACAS6456;
 XX
 DT 06-JUN-2003 (first entry)
 XX
 DE Human signalling pathway polynucleotide probe SEQ ID NO 1054.
 XX
 KW Human; probe; ss; array element; Parkinson's disease;
 KW signalling pathway population; cancer; adenocarcinoma; leukaemia;
 XX immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
 XX
 OS Homo sapiens.
 XX
 PN US6500938-B1.
 XX
 PD 31-DEC-2002.
 XX
 PF 30-JAN-1998; 98US-00016434.
 XX
 PR 30-JAN-1998; 98US-00016434.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Au-Young J, Seilhamer JT;
 XX
 DR MPI: 2003-352189/33.
 XX
 PT Combination of polynucleotide probes, useful as array elements in a
 PT microarray for monitoring the expression of a number of target
 PT polynucleotides.
 PS
 PS Claim 1; SEQ ID NO 1054; 65pp; English.
 XX
 CC The invention relates to a combination which, comprises a number of
 CC polynucleotide probes comprising a sequence selected from one of the 1490
 CC sequences mentioned in the specification. The combination is useful as an
 CC array element in a microarray for monitoring the expression of a number
 CC of target polynucleotides. The microarray is particularly useful in the
 CC diagnosis and treatment of cancer and immunopathology and neuropathology.
 CC The microarray is useful in diagnostics and treatment regimens, drug
 CC discovery and development, toxicological and carcinogenicity studies,
 CC forensics and pharmacogenomics. The microarray is also useful for
 CC monitoring progression of diseases and for developing therapeutic drugs. The
 CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs
 CC and genomic fragments and in research and diagnostic applications. The
 CC array can detect changes in expression in a large number of genes coding
 CC for different signalling pathway populations which can be used to diagnose
 CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
 CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
 CC and Parkinson's disease. The present sequence represents a polynucleotide
 CC probe of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=06500938B1
 CC
 SQ Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;
 Query Match 4.7%; Score 96.6; DB 7; Length 2320;
 Best Local Similarity 51.3%; Pred. No. 2.2e-08;
 Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 788 AAGTATATTGACATATGAGGCTGTGATTAAGCTATATGACCATATTATTAGA 847
DB 964 AGAGAAACCCCTACAAATGTAAAGATGTGTAAAGCTTTTAAACGATCTTCAACCCCTTAC 1023
QY 848 GCAACATTTAAGAACCCAGATATGATGACCGGTAAATGTACAGTGGACGATTGTGA 907
DB 1024 TACCATAGAAAAATTCATCTGAGAGAAACCTTACAAATGT-----GAAATATGTGG 1077
QY 908 TAAAGCATTTTGAAGAAATACATTTGGAACACATATTTGATTCATTCGAAAAAAA 967
DB 1078 CAAGCCTTTAAGACGCTCAAAACCTTACTACACATTAAGATTAATTCATCTGAGAGAA 1137
QY 968 ACCATTCATTTGCTGCTGTGTGTAAAGGGTTAATCTCCGACACATCTTGAAGACA 1027
DB 1138 ACCCTCAAAATGTAAAAATGTGAAAAAGCCTTTAACGAGCTGACACCTTACCAACA 1197
QY 1028 TGAATCACCCTATACAAAGTCATTTAATGTATGAAATTTGTCAAGACATTTTA 1087
DB 1198 TGAGTATATTCATACCTGAGAGAAACCCCTACAAATGTGAAAAATGTGAAAAAGCCTTTA 1257
QY 1088 TAAACA--TCAATCTTTAAGACATCATATATATCTGTTCAATGAAAAAATTAACGTG 1144
DB 1258 TCAATTTCTCACACCTTACTACATTAAGATTAATTCATGAGAGAAACCTTACAAATG 1317
QY 1145 TAAACATGTAATAAGTTTTCATGCTGACCTTCAAAATTTAGCAACATTAATTAACA 1204
DB 1318 TAAAGATGTGTAAAGCTTTTAAACCTTTCAACCTTACTTAACATTAAGTAAATTTCA 1377
QY 1205 TCATGTGATCTCTGCTTATCATGTGATCATCTGTTGTTTAAAAATTTCCAAAC 1264
DB 1378 TACTGAGAGAGAGCTTACAAATCTAAGATGTGAAAAAGCTTTTAACCAATCTGAAA 1437
QY 1265 TTGGTCAGATTTACATTTATATTAATAACAATGCAATCCAAAATTAATGTCTTAATG 1324
DB 1438 ACTTACTGAACATTAAGAAAAATTCATCTGAGAGAAACCCCTA--TGAATGTGAAAAATG 1494
QY 1325 TGTGTAAGGTGTGTGGGAAAAAGTTATCTGTCATATGTTAGTATGATGTA 1381
DB 1495 TGGCAAGCTTTTACCAAGTCTCAATCTTACTAGCATTAAGAAAGTCAATACAGA 1551

RESULT 6
AAS91317
ID AAS91317 standard; cDNA; 4563 BP.
XX
AC AAS91317;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27121.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
PA (HYSB-) HYSBQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR P-PSDB; ABG27130.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27121; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probe, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;
Query Match 4.6%; Score 95.2; DB 5; Length 4563;
Best Local Similarity 52.7%; Pred. No. 4.2e-08;
Matches 256; Conservative 0; Mismatches 223; Indels 9; Gaps 2;
QY 726 GAAAGTGACAAACCAANTGATATCATCTTTAAATCTCTCTTCTATCACTGCC 785
DB 1315 GAATGTGGCAAAAGCTTTAACCAGTCTCCACCTTACTCAATAAACAAATTCATACT 1374
QY 786 AAAAAGTATTTTGCATATGAAAGGTGTGAATAGCCATATAATGACCATCTTTTA 845
DB 1375 GAGGGAACCTTACAAATGTGAAGATGTGGCAAAAGCTTTTAAACATCTTGAGCCCTT 1434
QY 846 GAGCAACATTTAAGAACCCAGTAAATGATGACCGGTAAATGTACAGTGGACGATTGT 905
DB 1435 ACTMAACATTAAGTATTTCTACTGGGAGAGAAACCAATCAAAATGT-----GAAGATGT 1488
QY 906 GATTAAGCATTTTTCAGAAAAATGACATTGGAACACATATTTGATACATTTCCGAAAA 965
DB 1489 GGCAAAGCTTTTGGCAATCTCACAACCTTACTAGACATTAAGCAATTCATCTGGAGAG 1548
QY 966 AAACCTTCATTTGCTGATGTGTGTGAAGGGGTAAATCTGACAAACCTTGAAGA 1025
DB 1549 AAACCTTACAAATGTGAAGATGTGGCAAAAGCTTTTAAACATTTCTGAGACCTTGAAGA 1608
QY 1026 CATGAATCACCCATCAAAAGTCAATTTAAATGTACATTTGAAATTTGTCAAGAGCATTT 1085
DB 1609 CATAGATTAATTCATCTGAGAAAAACCTTACAAATGTGAAGATGTGGAAAAAGCTTTT 1668
QY 1086 TATTAACATCAATCTTTAAGA--CATCATATATTTATCTGTTCAATGAAAAACATTAACG 1142
DB 1669 AGCCAGTCTCAACCCCTTAAAGAACATCATATTAATTCATCTGAGAGAAACCCCTAACAA 1728
QY 1143 TGTAAACAATGTAATAAGTTTCACTGACCTTCAAAATTTAGACAAACATTAATTAATA 1202
DB 1729 TGTGAAGAAATGTGTAAAGCTTTTAAAGTGTCAATCAAAACTTACTGTACATTAAGTAAAT 1788
QY 1203 CATCATGCTG 1212
DB 1789 CATACTGGAG 1798

RESULT 7

AA151569
ID AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN MO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
XX
PR 29-JUN-2001; 2001US-0301787P.
XX
PR 29-JUN-2001; 2001US-0301792P.
XX
PR 29-JUN-2001; 2001US-0301892P.
XX
PR 06-JUL-2001; 2001US-0303405P.
XX
PR 06-JUL-2001; 2001US-0303442P.
XX
PR 15-MAR-2002; 2002US-0364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AR, Warren BA, Emerling BW,
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe JD;
PI Ramkumar J, Griffin JA, Yang J, Sanjwalala MM, Baughn MR;
PI Borowsky MI, Yao MG, Walla NK, Bandman O, Lal PG, Becha SD, Lee SY;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjian Y, Lu Y;
XX
DR WPI; 2003-201420/19.
XX
DR P-PSDB; AAO16419.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS).
XX
PS Claim 12; Page 292-293; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC the invention are useful for creating transgenic animals to model human
CC disease. The present DNA sequence encodes a human nucleic acid-associated
CC protein of the invention
XX
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;
XX
Query March 4.6%; Score 94.6; DB 7; Length 2597;
Best Local Similarity 55.1%; Pred. No. 5.1e-06;
Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;
XX
QY 797 TTGCACATATGAAAGGTGTGATTAAGCCTATATCAACCATTTATTAAGCAACATTT 856
DB 1070 TTATTAATGTGAAGAAATGTGGCAAAAGCCTTTAATGTGTCTCTCAACCTTACTCAACTAA 1129

QY 857 AAGAACCCACAGATATGATGACCCGTAATAATGTACAGTGGACGATTGTATTAAGCATTT 916
DB 1130 GAGAAATTCATCTCGAGAGAAACCTTACAAATGT-----GAAGAGTGTGGCAAAAGCTTT 1183
QY 917 TTTCAAGAAATTCATCTTTGAAACACATATTTGTATGACATTCGGAATAAAACCATTTCCA 976
DB 1184 TTAACGTCTCTCAACCTTACTACATTAAGAAATTTCTACTGAGAAACCATTAACAA 1243
QY 977 TTGTTCAAGTGTGTGTAAAGGGGTAAATTTCTGACAAACCTTGAAGAAACATGAAATCAC 1036
DB 1244 ATGTGAAGATGTGGCAAAAGCCTTTAACAATCTCTCAACCTTCAACCATTAAGAAAT 1303
QY 1037 CCATTCGAAGTCAATTAATGTATCATTTGAAATTTGTCAAGAACATTTTATTA---ACA 1093
DB 1304 TCAATCCGAGAGAAACCTTACAAATGTGAAGAAATGTGCAAAAGCCTTTAACAAGTTCTC 1363
QY 1094 TCAATCTTTAAGACATCATATATTAATCTGTCATGAAAAACATTAAGGTAAACAAAG 1153
DB 1364 ACACTTCTACATCAATTAAGTATTCATCTGAGAGAAACCTTACAAATGTAAAGAAATG 1423
QY 1154 TTAATTAAGTTTCTACTCGACCTTCAAAATTTAGCAACATTAATTAACATCATGTTG 1212
DB 1424 TGGCAAAAGCTTTAAGCGGTCTCAAACTTACTGAACTAGATTAATTCATCTGAG 1482
RESULT 8
AAS64586
ID AAS64586 standard; cDNA; 3639 BP.
XX
AC AAS64586;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #390.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
XX 11-OCT-2001.
XX
PD 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dumanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG00399.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 390; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food

supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. A864197-A864564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

Query Match 4.5%; Score 92.4; DB 5; Length 3639;

Best Local Similarity 54.1%; Pred. No. 1.3e-07;

Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

```

QY 788 AAGTATATTGACATATGAAAGGCTGATTAAGCCTTAATCCGACATCTATTATGA 847
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 AAGAGAAACCCCTCAATGTAAGAAATGCGCAAGCTTTAGCCATTCTTCAACCTTGC 797
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 848 GCACACTTAAGAACCCACAGTAATGATGACCGGTAAATGTACAGTGCAGATTGTA 907
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 TAAACATTAAGAGATTCATATCTGAGAGAAACCCCTCAAAATGT-----GAAGAAATGTG 851
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 908 TAAAGCATTTTTCAGAAATACATTTGGAACACATATTGTATCATCTCCGAAAAA 967
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 CAAGCTTTTAGCCGTTCTTCAACCTCTGTAACATTAAGAAATTCATATCTGAGAGAA 911
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 968 ACCATTCCATTGTCAGTGTGTGTAAAGGGTATTCTCCGACACACTGAAAAAGA 1027
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 ACCCTCAAAATGTAAGAAATGTGGCAAGCTTTTACCAATCTCAACCTTCTATATCA 971
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1028 TGAATACCCCATACAAAGCATTTTAATGTATGTAATTTGAAATTTGCAAGAACATTTTA 1087
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972 TAGAGTACTCATCTGAGAGAAACCCCTCAAAATGTAAGATGTGACAAACCTTTAA 1031
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1088 TAAACATCAATC---TTTAAACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 1144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1032 GCGACCTCTCAACCTTACTTAACATTAATAATATACATGCTGAGAGAAACTCTACAAATG 1091
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1145 TAAACATGTATTAAGTTTTCATCTGACCTTCAAAATTTAGCAACATTAATTTAAACG 1204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1092 TGAAGATGTGGCAAGCTTTTATGATCTTCAANTCTTATATACATTAAGTTTATTC 1151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1205 TCATGGTGAATCTCT 1220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 TACTGAGAGAAACT 1167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

ADB63624

ID ADB63624 standard; cDNA; 2110 BP.

XX ADB63624;

XX 04-DEC-2003 (first entry)

DE Human cDNA encoding clone THYMU20071120.

XX Human; sex; gene; pharmaceutical; diagnostic; gene therapy;

KW tissue regeneration; cell regeneration; membrane protein;

KW signal transduction-related protein; transcription-related protein;

KW osteoporosis; neurological disease; cancer; tumour.

XX Homo sapiens.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 170..1792

FT /tag= a

FT /product= "Clone THYMU20071120 protein"

XX EPI308459-A2.
 PN
 XX 07-MAY-2003.
 PD
 XX 28-MAR-2002; 2002EP-00007401.
 PF
 XX 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX WPI; 2003-450961/43.
 DR P-PDB; ADB65594.

New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.

Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesizing the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

XX Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

Query Match 4.5%; Score 91.8; DB 9; Length 2110;

Best Local Similarity 51.7%; Pred. No. 1.6e-07;

Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;

```

QY 780 CGTCCCAAAAAGTATATTGACATATGAAGGCTGTGATTAAGCCTTAATTCGACATCA 839
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 CATACGAGAGAAACCCCTCAAAATGTAAGAAATGTGGCAAGCTTTTAAACACTCTCA 1312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 840 TTATTAGAGCAACATTTTAAGAACCCACAGTAATGATGACCGGTAAATGATACAGTGCAC 899
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1313 GCCCTTACTCACTAATAGAGATTTCACTGAGAGAAACCCCTACAAATGT-----GAA 1366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 900 GATTGTATTAAGCATTTTTCAGAAATACATTTTGGAAACACATATTGTATCACTTCC 959
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1367 GAATGTGGCAAGCTTTTAAACGATCTCTCAAACTTACTGAACATTAAGAACTTCATAC 1426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 960 GAAAAAAACCATTCATTTGTCAGTGTGTGTAAAGGGGTAAATTTCTCGAACAACCTTG 1019
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1427 GAAAGAAACCCCTCAAAATGTAAGAAATGTGGCAAGCTTTTATCAACATCTCAAACTT 1486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1020 AAAAGCATGAATACCCCATCAAAAGCATTTTAAATCTATCATTTGAAAAATTTGTCAAGA 1079
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masudo Y;
 XX
 XX MPI; 2003-450961/43.
 DR P-PSDB; ADB64438.
 XX
 XX New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 PS
 PS Claim 1; Page; 222pp; English.
 XX
 XX The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence
 CC data for this patent is not represented in the printed specification, but
 CC is based on sequence information supplied by the European Patent Office.
 CC
 XX
 XX Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;
 SQ
 Query Match 4.5%; Score 91.8; DB 9; Length 3078;
 Best Local Similarity 52.1%; Pred. No. 1.7e-07;
 Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;
 QY 730 GTGACGAAACCAATCGATATCATCTTATATCTTCTTCATCAGTCCCAAA 789
 DB 650 GTGGCAATATCAATTTGATGCTTTCACATTAACCTACACATTAAGAAATTCATCTAG 709
 QY 790 AGTATATTGGACATATGAAGGGTGTGATTAAGCCCTTAATCGACATCATTTATTAAGC 849
 DB 710 AGTATCTTCAATGTGAAGATGTGTAAAGCTTAATCGCTTCAACCCCTTACTA 769
 QY 850 AACATTTAAGAACCCAGATATGATGACCGCTTAATATGACAGTGGAGCATGTGATA 909
 DB 770 AACATTAAGATTAATTCATCTGAGAAAAACCTTCAATGTG-----GAAGAATGTGCA 823
 QY 910 AAGCATTTTTCAGAAATTCATTTGGAACACATATTTGATCATTTCCGAAAAAAC 969
 DB 824 AAGCTTTTAACCGGTCTCTCAATCTTACTTAACATTAATTAATCTATCTGAGAGAAAC 883
 QY 970 CATTCATTTGTTCAAGTGTGTGAAGGGTTAATTCGACCAACACTTGAAGAAACATG 1029
 DB 884 CCTCAATGTGAAGAAATGTGGCAAGCTTTTAACCGGTCTCAACCTTACTTAACATA 943
 QY 1030 AATATCCCATCAAAAGTCATTTAAATGTACATTTGAAAAATTTGCAGAAACATTTTATA 1089
 DB 944 AAAAATTCATATCAAGAAAGAAACCCCTTCAATATGTGAAGATGTGGCAAGGCTTTAAC 1003

QY 1090 AACATCAATCTTT---AAGACATCATATATATCTGTTCTGATGAAAAAACATTAACGTGTA 1146
 DB 1004 AGTTCGATTTCTTATATTAAGCAATTAAGAAATTCATATGGAAGATTAACCTCAAAATGTG 1063
 QY 1147 AACAAATGTAATTAAGTTTCTACTGACCTTCAAAATAGCAACAAATTAATTAACATC 1206
 DB 1064 AAGAATGTGGCAAGCGCTTTAGAGTATCTCAATTTCTTAATAAACAATTAAGATTAATCCATA 1123
 QY 1207 ATGTGTGATCTCC 1219
 DB 1124 CTGGGAAAAACC 1136
 RESULT 12
 AAD55863
 ID AAD55863 standard; cDNA; 2026 BP.
 XX
 XX AAD55863;
 AC
 XX
 DT 07-AUG-2003 (first entry)
 XX
 DE Human nucleic acid associated protein (NNAP)-34 cDNA.
 XX
 XX Human; nucleic acid associated protein; NNAP; stroke; AIDS; neurologic;
 KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;
 KW developmental disorder; antiinflammatory; neuroprotective; thymomatous;
 KW Cushing's syndrome; infection; gene therapy; cyostatic; anticonvulsant;
 KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;
 KW gene; sr.
 XX
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 126..1973
 FT /tag= a
 FT /product= "human NNAP protein"
 XX
 PN MO200300648-A2.
 XX
 XX 23-JAN-2003.
 PD
 XX
 XX 10-JUL-2002; 2002MO-US021971.
 PF
 XX
 XX 12-JUL-2001; 2001US-0305089P.
 PR 12-JUL-2001; 2001US-0305104P.
 PR 13-JUL-2001; 2001US-0305325P.
 PR 13-JUL-2001; 2001US-0305390P.
 PR 19-JUL-2001; 2001US-0306960P.
 PR 20-JUL-2001; 2001US-0306949P.
 PR 27-JUL-2001; 2001US-0308170P.
 XX
 XX (INCYTE-) INCYTE GENOMICS INC.
 PA
 XX Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;
 PI Yue H, Baughn MK, Emerling BM, Lal PG, Lu DM, Forsythe J;
 PI Rankumar J, Li JX, Becha SD, Duggan BM, Sanjankwala MM, Lee EA,
 PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG,
 PI Barroso I, Tran B, Walla NK, Hafalia AJA, Nguyen DB, Lu Y;
 PI Arizizu CS;
 XX
 XX MPI; 2003-221732/21.
 DR P-PSDB; AAE37047.
 DR
 XX
 XX New human nucleic acid associated proteins (NNAP), useful for diagnosing,
 PT treating and preventing diseases or conditions associated with the
 PT aberrant NNAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
 PT infections.
 PT
 XX
 XX Claim 5; Page 259-260; 260pp; English.
 PS
 XX The invention relates to human nucleic acid associated proteins (NNAP)
 CC and their corresponding nucleic acid sequences. The invention is useful
 CC in diagnosing, treating and preventing diseases or conditions associated

CC with the decreased expression or overexpression of NAAp, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
 CC disorders, or infections. It is also useful in assessing the effects of
 CC exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of NAAp. The NAAp or its fragments are useful in screening
 CC compounds for which acts as their agonist or antagonist. The microarray
 CC is useful in monitoring or measuring protein-protein interactions, drug-
 CC target interactions, and gene expression profiles. NAAp DNA is used in
 CC gene therapy. The present sequence is human NAAp cDNA

XX Sequence 2026 BP; 699 A; 422 C; 396 G; 509 T; 0 U; 0 Other;

Query Match

Best Local Similarity 4.4%; Score 90.4; DB 7; Length 2026;
 Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY 726 GAAAGTGAAGAAACCAATCGATATCATCTTTAATCTCTTCTTCTTCACTGCTCC 785
 DB 1002 GAATGTGGAAAGCCCTTCAATCTTCACTTCAACCTTACACATTAAGGTAATTCATCT 1061
 QY 786 AAAAGTATATTTGACATATGAAGGTGTGAATTAAGCCTATATGACATCATATTA 845
 DB 1062 GGAGAGAAAGCCCTTCAATGTGAAGATGTGTAAAGCTTTAAACCACTTCAAGCCTT 1121
 QY 846 GAGCAACATTTAAGAACCCACAGTAATGATCGACGTATAATGTACAGTGAAGCATTTGT 905
 DB 1122 ACTACACATMAAGTTCATTCATGTATTAAGAAAAACCTCAAAATGT-----GAAAGATGT 1175
 QY 906 GATTAAGCATTTTTCAGAAATATCATCTTTGAAACCATATGTATCATATCCGAAAA 965
 DB 1176 GACAAAGCTTTTACCGATTCATCTTACTTAACATTAAGATATTCATTCCTGAGAG 1235
 QY 966 AAACATTCATTTGTTCAGTGTGTGTGAAGGGGTTAATTCGCAACACTTGAAGAA 1025
 DB 1236 AAATCTTCAATGTGAACAATGTGGCAAGGCTTTAATGTCTTCAACCTTCAAAA 1295
 QY 1026 CATGAATATCCCATACAAAGTCATTTAATGTATCAATTTGAAATTTGCAAGACATTT 1085
 DB 1296 CATGAAGATTTCAATCGAGAGAAACCTTCAAAATGTGAAGATGTGGCAAGCCTTT 1355
 QY 1086 TATAAATCAT---CAATCTTTAAGACATCATATATATCTGTTCAAGAAAAACATTAAG 1142
 DB 1356 AATGTCTTCAACCTTACTACATTAAGATGATTCATCTGAGAGAAACCTTACAAA 1415
 QY 1143 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCACAACATTAATTA 1202
 DB 1416 TGTGAAGATGTGGCAAGCCTTTTAACACATCCTCAAAATTAATTAATTAATTA 1475
 QY 1203 CATCATGTGTGATCTCCT 1220
 DB 1476 CATACTGGAGAGAAACCT 1493

RESULT 13

ADA53124
 ID ADA53124 standard; cDNA; 2114 BP.

XX ADA53124;

DT 20-NOV-2003 (first entry)

DE Human coding sequence, SEQ ID 692.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

KM inflammatory disease; osteoporosis; neurological disease; gene; ss.

OS Homo sapiens.
 XX EPI293569-A2.

PD 19-MAR-2003.

XX 21-MAR-2002; 2002BP-0006586.

PR 14-SEP-2001; 2001BP-00328381.

PR 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y,

DR MPI; 2003-395539/38.

XX P-PsDB; ADA54763.

PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.

PS Claim 1; SEQ ID NO 692; 205BP; English.

CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 2114 BP; 717 A; 439 C; 434 G; 524 T; 0 U; 0 Other;

Query Match

Best Local Similarity 4.4%; Score 90.4; DB 7; Length 2114;
 Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY 726 GAAAGTGAAGAAACCAATCGATATCATCTTTAATCTCTTCTTCTTCACTGCTCC 785
 DB 1024 GAATGTGGCAAGCCTTCAATCTTCACTTCAACCTTACACATTAAGGTAATTCATCT 1083
 QY 786 AAAAGTATATTTGACATATGAAGGTGTGAATTAAGCCTATATGTACAGCATATTA 845
 DB 1084 GAGAGAAAGCCCTTCAATGTGAAGATGTGTAAAGCTTTTAAACCACTTCAAGCCTT 1143
 QY 846 GAGCAACATTTAAGAACCCACAGTATGATGACCGTATAATGTACAGTGAAGATGT 905
 DB 1144 ACTACACATTAAGTTCATCTTCACTTAAAGAAAAACCTTCAAAATGT-----GAAAGATGT 1197
 QY 906 GATTAAGCATTTTTCAGAAATATCATCTTTGAAACCATATGTATCATATCCGAAAA 965
 DB 1198 GACAAAGCTTTTACCGATTCATCTTACTTAACATTAAGATATTCATCTTGGAGAG 1257
 QY 966 AAACATTCATTTGTTCAGTGTGTGTGAAGGGGTTAATTCGCAACACTTGAAGAA 1025
 DB 1258 AAATCTTCAATGTGAACAATGTGGCAAGGCTTTAATCTGATCTTCAACCTTCAAAA 1317
 QY 1026 CATGAATATCCCATACAAAGTCATTTAATGTATCATTTGAAATTTGCAAGAGCATTT 1085
 DB 1318 CATGAAGATTTCACTTACGAGAGAAACCTTCAAAATGTGAAGATGTGGCAAGCCTTT 1377
 QY 1086 TATAAATCAT---CAATCTTTAAGACATCATATATATCTGTTCAAGAAAAACATTAAG 1142
 DB 1378 AATGTCTTCAACCTTACTACATTAAGATGATTCATCTGAGAGAAACCTTACAAA 1437
 QY 1143 TGTAAACATGTATTAAGTTTCACTGACCTTCAAAATTAAGCACAACATTAATTA 1202
 DB 1438 TGTGAAGATGTGGCAAGCCTTTTAACACATCCTCAAAATTAATTAATTAATTA 1497
 QY 1203 CATCATGTGTGATCTCCT 1220
 DB 1498 CATACTGGAGAGAAACCT 1515

RESULT 14

ADAs52931
ID ADAs52931 standard; cDNA; 2230 BP.
XX AC
XX ADAs52931;
XX DT 20-NOV-2003 (first entry)
XX DE Human coding sequence, SEQ ID 499.
XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Neotropic;
XX KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX OS Homo sapiens.
XX PN EP1293569-A2.
XX PD 19-MAR-2003.
XX PF 21-MAR-2002; 2002EP-00006586.
XX PR 14-SEP-2001; 2001UP-00328381.
XX PR 24-JAN-2002; 2002US-0350435P.
XX PA (HELI-) HELIX RES INST.
XX PA (REMS-) RES ASSOC BIOTECHNOLOGY.
XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
XX PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX DR WPI. 2003-395539/38.
XX DR P-PSDB; ADAs54570.
XX PR
XX PT New polynucleotides encoding full-length polypeptides, e.g. secretory in
XX PT and/or membrane proteins, useful for developing medicines for diseases in
XX PT which the gene is involved, or as target molecules for gene therapy.
XX PS Claim 1; SEQ ID NO 499; 205bp; English.
XX CC The present invention relates to novel human secretory or membrane
XX CC proteins (ADAs54072-ADAs5710) and their coding sequences (ADAs52433-
XX CC ADAs54071). The coding sequences are useful in the gene therapy of
XX CC diseases caused by abnormalities of the proteins, e.g. cancer,
XX CC inflammatory diseases, osteoporosis or neurological disease.
XX SQ Sequence 2230 BP; 831 A; 434 C; 400 G; 565 T; 0 U; 0 Other;

Query Match 4.4%; Score 90.2; DB 7; Length 2230;
Best Local Similarity 50.1%; Pred. No. 3.2e-07;
Matches 338; Conservative 0; Mismatches 323; Indels 14; Gaps 4;

QY 710 ACTTATTAATGAGTGAAGTAGAGGAAACCAATGATATCATCTTAAATATCTTCTC 769
DB 640 ACGTTTCAGATGTGAAGATGTGACAAATCATCTTGATGCTTACGCCCTTAACA 699
QY 770 TTCTTCATCAGCTCCCAAAAGTATATTGGCATATGAGGGTGTGTAAGCTATTA 829
DB 700 TAAAAAATTCACTACTAGAGAAATTTCTACAAATGTGAAGTGTGAAAAAACCCTTAA 759
QY 830 TCGACCATCATTTATTAGACCAATTTAAGAACCAACAGTATATGACGGTAAATG 889
DB 760 CTGGTCACAAACCTTTCTTAAACCTTAAGAAATTTCTACTGAGAAAAAACCTTCAAAATG 819
QY 890 TACAGTGAAGCATTTGTGATTAAGCATTTTTCAGAAAATCAGATTGTGAAACATATTGT 949
DB 820 TGAAGT-----TGTGAAAAAGCCTTTCACCAATCTCTAATCCTTCAATAACATTAAGAT 873
QY 950 ATCAGATTTCGAAAAAATCCATTCATCTGTTCACTGTTGAGTGTGTAAGGGGTAAATCTTCG 1009
DB 874 AATTCTGACTGGAGAAAAAACCTTAAATGTGCACACTGTGGCAAGGCTTTAAACAGTTC 933
QY 1010 ACGAAGCTTGAAGAGACATGAATTCACCATATCAAGTAGTATTTAATGTACTTTGAAA 1069

Db	934	CTCACACCTTACTAGACATTAAGATTAATTCATCTAGGAGAGAAACCTTACAAATTTGTACCA	993
Qy	1070	TTTGTCAGAAAGCATTTTATATAAATCATCATCTTAA--AGACATCATATATATCTGTTC	1126
Db	994	ATGTGGCAAGGCTCTTTAAGCAGTCCCAACCTTACTTAACATATAGTATTTTACTGG	1053
Qy	1127	TGAATAAAACATTAACGTGTAAACATGTATATTAAGTTTCACTGACCTTCAAAATTAGC	1186
Db	1054	AGAGGAAACCATACAAATGTGTAGGAAATGTGGCAAGCTTTTAACCTATCTTAACCACTTAC	1113
Qy	1187	ACAACATTAATTAATAACATCATGTGTG-GATCTCTGCTTATCAATGTGATCATCTGTGTT	1245
Db	1114	TGAACATTAAGAAATTTAACACTTAAGAGAAAGCCTTACAAATGTGAAGATGTGGCAAGC	1173
Qy	1246	GTTTAAAAAATTTTCCAACTTGGTCAGATTTACAATTTCAATATAAACAATCGATCCAA	1305
Db	1174	CTTTAACCAGTTTTCACCCCTTATTTACACATTAAGTATTAATTCATGTGCAAGAGAAACCC-	1231
Qy	1306	AACCTTAATGTCTTAAGTGTGAAGGTGTGTGGGAAAAAGGTTATCTTCTCACATA	1365
Db	1232	--CAAAATGTGAAGAAATGTGGCAGAGGCTTTTAAACCATGTCCGCAAGCTCACTGACATTA	1285
Qy	1366	TGTTAAGTCATGATG 1380	
Db	1290	AGTTATATTCATACCTG 1304	

CC	ADCS8104
XX	ID ADCS8104 standard; cDNA; 1952 BP.
XX	
AC	ADC58104;
XX	
DT	18-DEC-2003 (first entry)
XX	
DE	Zinc finger protein 85-54.12 cDNA SEQ ID NO:1.
XX	
XW	ss; gene; zinc finger protein; cancer; haemopathy; HIV; immune disease.
OS	Unidentified.
XX	
FH	Key Location/Qualifiers
FT	CDS 173..1651
FT	/tag=a
FT	/product="zinc finger protein 85-54.12"
XX	
PN	CN1382706-A.
XX	
PD	04-DEC-2002.
XX	
Pf	26-APR-2001; 2001CN-00112739.
XX	
PR	26-APR-2001; 2001CN-00112739.
XX	
PA	(BLOW-) BLOWINDOW GENE DEV INC SHANGHAI.
XX	
PI	Mao Y, Xie Y;
XX	
DR	WPI; 2003-269470/27.
DR	P-PSTDB; ADCS8105.
XX	
PT	Polypeptide-zinc finger protein 85-54.12 and polynucleotide for coding it.
PS	
XX	Claim 6; SEQ ID NO 1; 33bp; Chinese.
XX	
CC	The invention relates to a novel zinc finger protein 85-54.12. The cc polynucleotide encoding the zinc finger protein, the application of the protein in treating diseases such as cancer, haemopathy, HIV infection and immune disease, the antagonist of the polypeptide and its medical action, and the application of the polynucleotide are also disclosed. The present sequence encodes the zinc finger protein of the invention.

XX Sequence 1952 BP; 760 A; 386 C; 321 G; 485 T; 0 U; 0 Other;

Query Match

Best Local Similarity 4.4%; Score 90; DB 9; Length 1952;
Pred. No. 3,4e-07;

Matches 267; Conservative 0; Mismatches 230; Indels 9; Gaps 3;

```
QY 710 ACTTTATTCATGAGGAAAGTGCAGAAACCAATGCATTCATCTTTAATATCTTCTTC 769
Db 469 ACCTTTAAATATTAAGAAATTTGGCAATCATTTTGCATATTTTCAAACTTAACCTCAACA 528
QY 770 TTCTTCATCAGTCCCAAAAAGTATTTGACATATGAAGGGGTGAATTAAGCTATAA 829
Db 529 TAAATAATTTTGACTAGAGTAAATTTCTACAAATGTGAAGACTGTGAAAAGCCTTTAA 588
QY 830 TCGACCATCATTTATTAAGCAACATTTAAGAACCAAGTAATGATCGACCGTATTAATG 889
Db 589 TGGATCCTCAATCTTACTTAACATTAAGAAATTTCAATTTGGAGAGAAATCGTACATATG 648
QY 890 TACAGTGAAGATTGTGATTAAGCATTTTCAGAAATCACATTTGAAAACATATTTGT 949
Db 649 T-----GAGAAATGTGGCAAAAGCCTGTAAACAGTTCACAAACTTACTACATPAAGAT 702
QY 950 ATCAATTCGAAAAAAAACCATTCATGTTCAGTGTGGTAAAGGGTTAATTCG 1009
Db 703 AATTATTACTAGAGCAAACTCTACAAAGTGAAGATGTAGCAAAAGCCTTAACTCTGC 762
QY 1010 ACACACTTGAAGAAGCATGAATCACCCATACAAAGTCATTAAATGTACATTTGAAAA 1069
Db 763 GTCAACATTCACACACATACATATATCATCTGAGAGAAATCCCTACAAACGTGAAGA 822
QY 1070 TTGTCAAGAGCATTTTAT-AAACATCATCTTTAAGATCATATATATCTGTTCATG 1128
Db 823 ATGTACAAAGCCTTTAAACAGTCCCTCAACCTTACTACACATPAAGATATTCATACAG 882
QY 1129 AAAAAACATTACG--TGTAACAATGTATTAAGTTTCACTCGACCTTCAAAATTAAG 1186
Db 883 AGAGAACTCAATGAATATTAAGAAATGTGGCAAGCTTTCAACAGTCCCGACACCTTAC 942
QY 1187 ACAACATPAATTTAAACATCATGTG 1212
Db 943 CAGACATPAAGATTAATTCATCTGAG 968
```

Search completed: July 28, 2004, 15:04:04
Job time : 828 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 12:51:57 ; Search time 8116 Seconds

(without alignments)
11001.312 Million cell updates/sec

Title: US-09-831-804-1

Sequence: 1

1 cttattatgaagatgctc.....gacatactactcttcatg 2060

Scoring table:

IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched:

3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:

6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2060	100.0	2060	6	BD274349
2	1239	60.1	1239	6	BD274350
3	1235.8	60.0	1239	6	AX489193
4	217.4	10.6	1560	8	YSCNOFPAT
5	217.4	10.6	1739	8	YSCFPITIA
6	217.4	10.6	37497	8	YSCP9677
7	199.2	9.7	2133	8	YSCRP026A
8	193	9.4	797	11	CNSO61NG
9	101	4.9	169063	9	AC138126
10	101	4.9	177299	9	AC073544
11	99.4	4.8	2132	6	AX714042
12	99.4	4.8	2132	6	AX056088
13	99.4	4.8	138627	9	AC099500
14	99.4	4.8	169500	2	AC138469
15	98.2	4.8	2237	9	AK122869
16	97.8	4.7	1203	9	D70831
17	97.8	4.7	235532	9	AC008739
18	97.2	4.7	115995	9	AC011494
19	96.6	4.7	2320	6	AR270491
20	96.6	4.7	2320	9	HS035376
21	96.6	4.7	68304	10	BX000432
22	95.2	4.6	1389	9	M27879
23	95	4.6	981	9	HS42F9
24	94.6	4.6	186233	9	AC092329
25	94.6	4.6	189317	2	AC024483
26	94.2	4.6	292390	2	AC105677
27	93.6	4.5	2873	9	HDMHPLK
28	93.6	4.5	2873	11	G28705
29	93.6	4.5	2873	11	HMSWS1269
30	93.6	4.5	155439	9	AC008626
31	93.6	4.5	191082	2	AC141066
32	93.6	4.5	199585	10	AC124426
33	93.6	4.5	203396	9	AC073210
34	92.4	4.5	41153	9	AC016628
35	92.4	4.5	158430	2	AC012431
36	92	4.5	156835	9	AC010620
37	91.8	4.5	1870	9	AY044432
38	91.8	4.5	2110	6	AX748253
39	91.8	4.5	2110	6	AX748253
40	91.8	4.5	2909	9	AK093669
41	91.8	4.5	2976	9	HSMB808021
42	91.8	4.5	3078	6	AX747097
43	91.8	4.5	3078	6	AK091618
44	91.8	4.5	3123	9	BC040594
45	91.8	4.5	174994	9	AC022145

ALIGNMENTS

RESULT 1
BD274349
LOCUS
DEFINITION
Candida albicans tflitA gene (CatfIttA) and the coded CATfIttA protein.
ACCESSION
BD274349
VERSION
BD274349.1 GI:33084117
KEYWORDS
JP 2002531068-A/1.
SOURCE
Candida albicans
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 2060)
Pallier, F.B., Camier, S. and Sentenac, A.
Candida albicans tflitA gene (CatfIttA) and the coded CATfIttA

JOURNAL
 Patent: JP 2002531068-A 1 24-SEP-2002;
 COMMENT
 AVENTIS PHARMA SA
 Candida albicans
 OS
 JP 2002531068-A/1
 PD 24-SEP-2002
 PF 09-NOV-1999 JP 2000581204
 PR 10-NOV-1998 FR 98/14147
 PI FLORENCE BORDON PALLIER, SYLVIE CAMIER, ANDRE SENTENAC PC
 C12N15/09, C12N15/09, A61K39/00, A61K45/00, A61P31/10, C07K14/40, PC
 C07K16/14,
 PC C12N1/19, C12N1/21, C12P21/02, C12Q1/02, G01N33/15, G01N33/50, PC
 G01N33/53,
 PC G01N33/566, G01N33/569// (C12N15/09, C12R1:725), C12N15/00, C12N15/00,
 PC 00,
 PC (C12N15/00, C12R1:725)
 CC Candida albicans efiligene (CatfIIIA) and the coded CatfIIIA
 CC protein
 FH Key
 FT source
 Location/Qualifiers
 1..2060
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /DB_xref="taxon:5476"

ORIGIN
 Query Match 100.0%; Score 2060; DB 6; Length 2060;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2060; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTTATTAGAGAGATTGGCTAGGCCCATTTTGTATACGGGCTCTCCAAAGTGCATTTGTTT 60
 DB 1 CTTTATTAGAGAGATTGGCTAGGCCCATTTTGTATACGGGCTCTCCAAAGTGCATTTGTTT 60
 QY 61 TACTAAATATCCCAATCATTTGGGCTCAGTGTGATGGGGGTGTGCATCTCTTGTTAG 120
 DB 61 TACTAAATATCCCAATCATTTGGGCTCAGTGTGATGGGGGTGTGCATCTCTTGTTAG 120
 QY 121 AATAGGCGGAGGCTCCGATCCCAAAAAAGAGATCAGATGTCTCGGCTGCAGA 180
 DB 121 AATAGGCGGAGGCTCCGATCCCAAAAAAGAGATCAGATGTCTCGGCTGCAGA 180
 QY 181 TTGTAGCCATGCGAAATGCCGAAAAATGAAAAAATGAAAAAATCTACGGGCCCACT 240
 DB 181 TTGTAGCCATGCGAAATGCCGAAAAATGAAAAAATGAAAAAATCTACGGGCCCACT 240
 QY 241 ACAAAAGAAAAGATTTGAATCTGATCAGTACAGTGTCTGGAACCTCTATATTTTAA 300
 DB 241 ACAAAAGAAAAGATTTGAATCTGATCAGTACAGTGTCTGGAACCTCTATATTTTAA 300
 QY 301 TATGTACAGGAGCTTTAGAAATTTGTATATTTGTGTCTGACACTGTGTGTTAAATCT 360
 DB 301 TATGTACAGGAGCTTTAGAAATTTGTATATTTGTGTCTGACACTGTGTGTTAAATCT 360
 QY 361 GGACATCTGTTCCCTTTGTGAAGGTCGTCTGTAATGAATTCATGATCAAGATTAATAT 420
 DB 361 GGACATCTGTTCCCTTTGTGAAGGTCGTCTGTAATGAATTCATGATCAAGATTAATAT 420
 QY 421 GACTTGTCTACCTCATAGATGCGGACTTGATTAATATTAGCTTTAATCCTCTGTATA 480
 DB 421 GACTTGTCTACCTCATAGATGCGGACTTGATTAATATTAGCTTTAATCCTCTGTATA 480
 QY 481 TATGTAAACCACTTGAATTTCTTGTGTGGGATTCACCTTGGATGATGATTTAC 540
 DB 481 TATGTAAACCACTTGAATTTCTTGTGTGGGATTCACCTTGGATGATGATTTAC 540
 QY 541 CAATGTATTTGTCATCTTTTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 DB 541 CAATGTATTTGTCATCTTTTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 QY 601 ATTTCATTAATGAGCGAGCGAATTCCTCTCTCATAGAAAAAATTAACAACCTGA 660
 DB 601 ATTTCATTAATGAGCGAGCGAATTCCTCTCTCATAGAAAAAATTAACAACCTGA 660

DB 601 ATTTCATTAATGAGCGAGCGAATTCCTCTCTCATAGAAAAAATTAACAACCTGA 660
 QY 661 AAAATTAACACCGGAGCTCATCTCTTTTCAAAATCACCTTTTATTAATTAATCA 720
 DB 661 AAAATTAACACCGGAGCTCATCTCTTTTCAAAATCACCTTTTATTAATTAATCA 720
 QY 721 TGAATGAAAAGTACGAAACCAATCGATATCATCTTTAATATCTTCTCTTCATCAG 780
 DB 721 TGAATGAAAAGTACGAAACCAATCGATATCATCTTTAATATCTTCTCTTCATCAG 780
 QY 781 GTCCCAAAAAGTATATTGGCAATATGAAAGGTCGTATGAAGCCCTATATGCAACATCA 840
 DB 781 GTCCCAAAAAGTATATTGGCAATATGAAAGGTCGTATGAAGCCCTATATGCAACATCA 840
 QY 841 TATTAGAGCAATTTTAAGACCCAGTAATGATGACCCGATTAATGATGAGTACAGTGAAG 900
 DB 841 TATTAGAGCAATTTTAAGACCCAGTAATGATGACCCGATTAATGATGAGTACAGTGAAG 900
 QY 901 ATTGTGATTAAGCAATTTTCAAGAAATCATTGGAAACATATTTGATGACATTCGG 960
 DB 901 ATTGTGATTAAGCAATTTTCAAGAAATCATTGGAAACATATTTGATGACATTCGG 960
 QY 961 AAAAAAACCATTCATTTGTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1020
 DB 961 AAAAAAACCATTCATTTGTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1020
 QY 1021 AAAGCATGAATTCACCCATCAAGTCAATTAATATGATCAATTTGAATAATGTCAGAGAG 1080
 DB 1021 AAAGCATGAATTCACCCATCAAGTCAATTAATATGATCAATTTGAATAATGTCAGAGAG 1080
 QY 1081 CATTTTATTAACATCATCTTTTAAGACATCATATATATCTGTGTGTGTGTGTGTGTGT 1140
 DB 1081 CATTTTATTAACATCATCTTTTAAGACATCATATATATCTGTGTGTGTGTGTGTGTGT 1140
 QY 1141 CGTGTAAACATGTAATTAAGTTTCACTGACCTCAAAATTAAGACAAATTAATTA 1200
 DB 1141 CGTGTAAACATGTAATTAAGTTTCACTGACCTCAAAATTAAGACAAATTAATTA 1200
 QY 1201 AACATCATGATGATCCGCTTATCAATGTATCATCTGTGTGTGTGTGTGTGTGTGTGT 1260
 DB 1201 AACATCATGATGATCCGCTTATCAATGTATCATCTGTGTGTGTGTGTGTGTGTGTGT 1260
 QY 1261 AAACCTGTCAATTAATCAATTTCAATTAATAACAATGATCCATTAATTTCTTA 1320
 DB 1261 AAACCTGTCAATTAATCAATTTCAATTAATAACAATGATCCATTAATTTCTTA 1320
 QY 1321 AATGTGTAAAGTTGT 1380
 DB 1321 AATGTGTAAAGTTGT 1380
 QY 1381 ATTCTACATGATCAAAATATGACCTTGATTAATGTGTGTGTGTGTGTGTGTGTGTGT 1440
 DB 1381 ATTCTACATGATCAAAATATGACCTTGATTAATGTGTGTGTGTGTGTGTGTGTGTGT 1440
 QY 1441 AAAATGAATTAAGTGAACATTTAATATCTTCAATGATGTAATCTCTGATGATTTAT 1500
 DB 1441 AAAATGAATTAAGTGAACATTTAATATCTTCAATGATGTAATCTCTGATGATTTAT 1500
 QY 1501 TAAAGGAAATGTAAGTGAATAATTAAGAAACCTATTAATGATCAAGATGCAATTAATA 1560
 DB 1501 TAAAGGAAATGTAAGTGAATAATTAAGAAACCTATTAATGATCAAGATGCAATTAATA 1560
 QY 1561 ATTGTGATTAAGTGAAGAAATTAAGAAATTAAGTGAAGAAATGAAGATGAAGAG 1620
 DB 1561 ATTGTGATTAAGTGAAGAAATTAAGAAATTAAGTGAAGAAATGAAGATGAAGAG 1620
 QY 1621 ATAGCTAGATGAAGAAAGATGATGATTAATGATGATGATGATGATGATGATGATGAT 1680
 DB 1621 ATAGCTAGATGAAGAAAGATGATGATTAATGATGATGATGATGATGATGATGATGAT 1680
 QY 1681 TAAATCATTTACTGCTCTTGTGAAGGTCGAAGAGGTTCTTAACCTTAATTTGAAATA 1740
 DB 1681 TAAATCATTTACTGCTCTTGTGAAGGTCGAAGAGGTTCTTAACCTTAATTTGAAATA 1740

Db	ATGAGTGAAGAAGGAGAAAACCAATCGATATCATCTTTAATATCTTCTTCTTCCTTCATCA	60
QY	780 CGTCCCAAAAAAGTATATTTGGCACATATGAAAGGGTGTGATAAAGCCTATATATGACCATCA	839
Db	61 CGTCCCAAAAAAGTATATTTGGCACATATGAAAGGGTGTGATAAAGCCTATATATGACCATCA	120
QY	840 TTATTAGAGCAACATTTAAGAAACCCACGTAATGATCGACCGTAAATGTACAGTGCAC	899
Db	121 TTATTAGAGCAACATTTAAGAAACCCACGTAATGATCGACCGTAAATGTACAGTGCAC	180
QY	900 GATGTGTGATAAAGCAATTTTTCGAGAAATACATTTGGAAACACATATGTATCAATATCC	959
Db	101 GATGTGTGATAAAGCAATTTTTCGAGAAATACATTTGGAAACACATATGTATCAATATCC	240
QY	960 GAAAAAACCATTCCATTTGTTCAAGTGTGTGGTAAAGGGGTAAATTTTCGACAAACATTTG	1011
Db	241 GAAAAAACCATTCCATTTGTTCAAGTGTGTGGTAAAGGGGTAAATTTTCGACAAACATTTG	300
QY	1020 AAAAGACATGAATATCAACCAATACAAAGTCATTTAAATGTATGAAATTTGTCAGAA	1079
Db	301 AAAAGACATGAATATCAACCAATACAAAGTCATTTAAATGTATGAAATTTGTCAGAA	360
QY	1080 GCATTTTATTAACATCAATCTTTAAGATCATATATATATCTGTTCAAGAAAAACATTA	1133
Db	361 GCATTTTATTAACATCAATCTTTAAGATCATATATATATCTGTTCAAGAAAAACATTA	420
QY	1140 ACGGTAAACAATGTAAATMAAGTTTCACTGCACCTTCAAAATTTAGACAACATTAATTA	1199
Db	421 ACGGTAAACAATGTAAATMAAGTTTCACTGCACCTTCAAAATTTAGACAACATTAATTA	480
QY	1200 AAAACATCATGGTGATCTCTGCTTATCATGTGATCATCTGCTGTGTTTAAAAATTC	1259
Db	481 AAAACATCATGGTGATCTCTGCTTATCATGTGATCATCTGCTGTGTTTAAAAATTC	540
QY	1260 CAAACTTGGTGAATTAACAATTCATATTAACAAATGCAATCCAAACTTAATGTCTCT	1319
Db	541 CAAACTTGGTGAATTAACAATTCATATTAACAAATGCAATCCAAACTTAATGTCTCT	600
QY	1320 AAATGTGTAAAGGTTGTGTGGAAAAAAGTTATCTTCAACATATGTAAAGTCATGAT	1379
Db	601 AAATGTGTAAAGGTTGTGTGGAAAAAAGTTATCTTCAACATATGTAAAGTCATGAT	660
QY	1380 GATTTCAACCATGATCAAAATATGGAATTTGTGATTTGTGATGTGGGAAATTTGCAAG	1439
Db	661 GATTTCAACCATGATCAAAATATGGAATTTGTGATTTGTGATGTGGGAAATTTGCAAG	720
QY	1440 AAAAAATATTTAGTGAACAATTAATATCTTCATATATGTAATATCCCTGATGATTTA	1499
Db	721 AAAAAATATATTTAGTGAACAATTAATATCTTCATATATGTAATATCCCTGATGATTTA	780
QY	1500 TTAAAGGAAACCTGAAGTGAAGAAAAATTTAGAAACCTATTAATCAAGAGTGAATTAAT	1559
Db	781 TTAAAGGAAACCTGAAGTGAAGAAAAATTTAGAAACCTATTAATCAAGAGTGAATTAAT	840
QY	1560 AATTTGATGAATTTAGAAACAGGAAATTAAGGTGAAGAAAGATGAAGAAATGAAGA	1619
Db	841 AATTTGATGAATTTAGAAACAGGAAATTAAGGTGAAGAAAGATGAAGAAATGAAGA	900
QY	1620 GATAGTCAATGAAAAAAGAAAGATGTTAAGTCAAGCTCAATGTCAAGTCAAAAGATCA	1679
Db	901 GATAGTCAATGAAAAAAGAAAGATGTTAAGTCAAGCTCAATGTCAAGTCAAAAGATCA	960
QY	1680 ATAAAAATCAATTAATCTGCTTCTTTGGAAAGGTTCAAGAGTGTCTTAAACTTAATCTGAAT	1739
Db	961 ATAAAAATCAATTAATCTGCTTCTTTGGAAAGGTTCAAGAGTGTCTTAAACTTAATCTGAAT	1020
QY	1740 AATGGGAAAGAAAGATCAATTTGCTTAAGATTAATATTTGTGATAGAAATGTTTCTAGAAATAT	1799
Db	1021 AATGGGAAAGAAAGATCAATTTGCTTAAGATTAATATTTGTGATAGAAATGTTTCTAGAAATAT	1080
QY	1800 GATTTACGTGCACATTTGAAATAGGACATGATGATTAATTTCAAAAGAAATGATGATCTTCTTA	1859
Db	1081 GATTTACGTGCACATTTGAAATAGGACATGATGATTAATTTCAAAAGAAATGATGATCTTCTTA	1140

QY 1860 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCTTGGTTAAAGAAAGCCAGATG 1913
 DB 1141 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCTTGGTTAAAGAAAGCCAGATG 1200
 QY 1920 GATTATGTCGAAATGAAACATCAGTGAATTCCTCGATTA 1958
 DB 1201 GATTATGTCGAAATGAAACATCAGTGAATTCCTCGATTA 1239
 RESULT 3
 LOCUS AX489193 1239 bp DNA linear PAT 16-AUG-2002
 DEFINITION Sequence 6493 from Patent WO02053728.
 ACCESSION AX489193
 VERSION AX489193.1 GI:22323205
 KEYWORDS
 SOURCE Candida albicans
 ORGANISM Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Microsporid Saccharomycetales; Candida.
 REFERENCE 1
 AUTHORS Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K.L.
 TITLE Gene disruption methodologies for drug target discovery
 JOURNAL Patent: WO 02053728-A 6493 11-JUL-2002;
 Elitza Pharmaceuticals, Inc. (US)
 FEATURES
 source 1..1239
 Location/Qualifiers
 /organism="Candida albicans"
 /mol_type="unassigned DNA"
 /db_xref="taxon:5476"
 ORIGIN
 Query Match 60.0%; Score 1235.8; DB 6; Length 1239;
 Best Local Similarity 99.8%; Pred. No. 8.2e-187;
 Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 720 ATGAGTGAAGAGTGAACGAAACCAATCGATATCATCTTTAATCTTCTCTTCATCA 779
 DB 1 ATGAGTGAAGAGTGAACGAAACCAATCGATATCATCTTTAATCTTCTCTTCATCA 60
 QY 780 CGTCCCAAAAGATATTTTGCACATATGAAGGCTGTAAAGCCCTAATTCAGCATCA 839
 DB 61 CGTCCCAAAAGATATTTTGCACATATGAAGGCTGTAAAGCCCTAATTCAGCATCA 120
 QY 840 TTATTAGAGCAACATTTAAGAACCCACAGTATGATGACCGTAAATGACAGTGAC 899
 DB 121 TTATTAGAGCAACATTTAAGAACCCACAGTATGATGACCGTAAATGACAGTGAC 180
 QY 900 GATTGTGATTAAGCATTTTTCAGAAATTCATTTGGAAACACATATTTGATCATTTCC 959
 DB 181 GATTGTGATTAAGCATTTTTCAGAAATTCATTTGGAAACACATATTTGATCATTTCC 240
 QY 960 GAAAAAAGCAATTCATTTGTCAGTGTGTAAAGGGTTAAATTCGCAACACATTG 1019
 DB 241 GAAAAAAGCAATTCATTTGTCAGTGTGTAAAGGGTTAAATTCGCAACACATTG 300
 QY 1020 AAAAGACATGAATACCCCATCAAAAGTCAATTTAAATGTACATTTGAAAAATTTGCAAGAA 1079
 DB 301 AAAAGACATGAATACCCCATCAAAAGTCAATTTAAATGTACATTTGAAAAATTTGCAAGAA 360
 QY 1080 GCATTTTATAAATCATCTTTAAGACATCATATATATCTGTTTCATGAAAAACATTA 1139
 DB 361 GCATTTTATAAATCATCTTTAAGACATCATATATATCTGTTTCATGAAAAACATTA 420
 QY 1140 ACGGTGTAACATGATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 1199
 DB 421 ACGGTGTAACATGATTAAGTTTTCATCGACCTTCAAAATTAAGCAACATTAATTA 480
 QY 1200 AAACATCATGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTGTTTAAAAATTC 1259
 DB 481 AAACATCATGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTGTTTAAAAATTC 540

QY 1260 CAAACTTGTCAGATTAATCAATTTCAATATAAACAATCGATCCAAACTTAATCTCT 1319
 DB 541 CAAACTTGTCAGATTAATCAATTTCAATATAAACAATCGATCCAAACTTAATCTCT 600
 QY 1320 AATGTGTAAAGGTTGTGTGGAAAAAGTTTATCTTCCACATGTGTTAAGTCAATG 1379
 DB 601 AATGTGTAAAGGTTGTGTGGAAAAAGTTTATCTTCCACATGTGTTAAGTCAATG 660
 QY 1380 GATTCACCATGATCAAAATATGAACTTGTGATTAATTTGATGTGGGAAATTTGCAAG 1439
 DB 661 GATTCACCATGATCAAAATATGAACTTGTGATTAATTTGATGTGGGAAATTTGCAAG 720
 QY 1440 AAAAAATGAATTAAGTGAACATTAATATCTTCATGATGTAATATCCCTGATGATTA 1499
 DB 721 AAAAAATGAATTAAGTGAACATTAATATCTTCATGATGTAATATCCCTGATGATTA 780
 QY 1500 TTAAAGAAACCTGAAGTGAAGAAATTTAGAAACCTTTTGAATCAAGATGGAATTAAT 1559
 DB 781 TTAAAGAAACCTGAAGTGAAGAAATTTAGAAACCTTTTGAATCAAGATGGAATTAAT 840
 QY 1560 AATTCGATGAATTAAGAAACAGAAATTTAAAGTGAAGAAAGATGAAGAA 1619
 DB 841 AATTCGATGAATTAAGAAACAGAAATTTAAAGTGAAGAAAGATGAAGAA 900
 QY 1620 GATAGCTTGAATGAAAAAGAAAGTGAATGATGATCACTCAATGTCTCAAAATCA 1679
 DB 901 GATAGCTTGAATGAAAAAGAAAGTGAATGATGATCACTCAATGTCTCAAAATCA 960
 QY 1680 ATTAATTCATTTTCTGCTCTTTTGAAGGTGAAGAGTTCCTCAACTTATCTGAAT 1739
 DB 961 ATTAATTCATTTTCTGCTCTTTTGAAGGTGAAGAGTTCCTCAACTTATCTGAAT 1020
 QY 1740 AGTGGGAGAAAGATCAATTTGCTTAAGAAATATTTGATGAATGTTTCTTAAGAAAT 1799
 DB 1021 AGTGGGAGAAAGATCAATTTGCTTAAGAAATATTTGATGAATGTTTCTTAAGAAAT 1080
 QY 1800 GATTTCAGTGAATTTGAAATGCAATGATGATTAATTTACAAAGATTTGCTTCTTA 1859
 DB 1081 GATTTCAGTGAATTTGAAATGCAATGATGATTAATTTACAAAGATTTGCTTCTTA 1140
 QY 1860 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCTTGGTTAAAGAAAGCCAGATG 1919
 DB 1141 AATAGTATGAAAGAAAGAAAGAACTCCAGAAAGGTGAACCTTGGTTAAAGAAAGCCAGATG 1200
 QY 1920 GATTATGTCGAAATGAAACATCAGTGAATTCCTCGATTA 1958
 DB 1201 GATTATGTCGAAATGAAACATCAGTGAATTCCTCGATTA 1239
 RESULT 4
 YSCNOFEAT 1560 bp DNA linear PLN 27-APR-1993
 LOCUS Saccharomyces cerevisiae transcription factor IIRA and RNA
 DEFINITION polymerase subunit RPB6 gene.
 ACCESSION M90638.1 GI:172046
 VERSION RNA polymerase; transcription factor IIRA.
 KEYWORDS Saccharomyces cerevisiae (baker's yeast)
 SOURCE Saccharomyces cerevisiae
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetales; Saccharomycetes.
 REFERENCE 1
 AUTHORS Moychik, N.A. and Young, R.A.
 TITLE Genes encoding transcription factor IIRA and the RNA polymerase
 common subunit RPB6 are divergently transcribed in Saccharomyces
 cerevisiae
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)
 MEDLINE 92237295
 PUBMED 1570325
 COMMENT Original source text: Saccharomyces cerevisiae (strain S288C) DNA.
 FEATURES
 source 1..1560
 Location/Qualifiers
 /organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"
/strain="S288C"
/db_xref="taxon:4932"
ORIGIN
chromosome 14.

Query Match 10.6%; Score 217.4; DB 8; Length 1560;
Best Local Similarity 59.4%; Pred. No. 4.4e-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

QY 781 GTCCAAAAGATATATTTGACATATGAAAGGGTGTGATTAAGCCCTATATGACCATCAT 840
DB 368 GGGCAAGACATATTTCTGTACTATGATGGCTGTGATTAAGGCAATTTACAGACCTTCAA 427
QY 841 TATTAGAGCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAAG 900
DB 428 TTTGACTGAACACCAATTAAGCGTACATCAGGGCTTAA--AGAGCATTTTCAGTGATATA 484
QY 901 ATTGTGATAAGCAATTTTTCAGAAAATCACATTTTGGAAACACATATGTATACATTTCCG 960
DB 485 AGTGTGCAAAATCTCTTGTTAAAAAGAGTCACTTAGAGAGCACTTGTATACGCAATCTG 544
QY 961 AAAAAAACCATTCCATTTGTCAGTGTGTGTAAGGGGTTAAATTCGACAAACACTTGA 1020
DB 545 ATACGAACCATTCACATGTTCTTATTTGTGAAAAGAGTGAAGSACTGCCAGCACTGA 604
QY 1021 AAGACATGAATACACCATACAAAGTCATTTTAATGTATACATTTGAAATTTGCAAGAG 1080
DB 605 AGGACACGAGTAGGACATACCAATCTTTCATTTGTCCAGAGGAAGATGCAACCTCC 664
QY 1081 CATTTTATAAACATCAATCTTTTAAGACATCATATATATCTGTCATGAAAAACATTTAA 1140
DB 665 GATTTACAAAGATCCACATTTAAGGACATATTTTATCTGTTCAATTTACATTAACATA 1200
QY 1141 CGTGTAACCAATTAATAAGTTTTCACCTGACCTTCAAAATTAAGACACATTAATA 1260
DB 725 CCTGTCCACACTGCATAAAGCTTTTCAGAGCCATATAGGCTTAAGAAATCAATTTCTA 784
QY 1201 AACATCA-----TGTGTGATCTTCCTTATCAATGTATCATCTGGTTGTTTAAAA 1264
DB 785 AACATCAAGATCTGTAGGTAGAAAATCTTACCAATGTACTTTGTGGTTGTGCAAG 844
QY 1255 ATTTCCAAACCTTGATGATTTACAAATTTCAATTAACAACTGCAACAACTTAAT 1314
DB 845 AGTTTGGATATGTGTCAATTTGCAATCGATTTAAAAATGATCATCTCAAAATTAATAAT 904
QY 1315 GTCCCTAATGTGTGTAAGGTTGTGTTGGAAAAAGTTATCTTCACATATGTATAGTC 1374
DB 905 GTCCCTAATGTGTGTAAGGTTGTGTTGGAAAAAGTTATCTTCACATATGTATATTC 964
QY 1375 ATGATGATCTACACATGATCAAAATATGACATTTGTGATTTGTGATGTGGGAAATTTG 1434
DB 965 ATGACGACTACATGATACCAAAATTTGAAAGTGTCAATATGTCTCGATATGTCTTTT 1024
QY 1435 CAAGAAAAATGATATGTAACATTTAT 1463
DB 1025 CTAGAAAACATGATCTTCTACGCAATTA 1053

RESULT 5
YSCFIIIA 1739 bp DNA linear PLN 13-SRP-1996
LOCUS Saccharomyces cerevisiae transcription factor IIfa (TFIIfa) gene,
DEFINITION complete cds.
ACCESSION M80611.1 GI:172902
VERSION M80611.1
KEYWORDS transcription factor IIfa; zinc-finger protein; zinc-finger
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 1739)
AUTHORS Archambault,J., Milne,C.A., Schapert,K.T., Baum,B., Friesen,J.D.

TITLE and Segall,J.D.
The deduced sequence of the transcription factor TFIIfa from
Saccharomyces cerevisiae reveals extensive divergence from Xenopus
TFIIfa
JOURNAL U. Biol. Chem. 267 (5), 3282-3288 (1992)
MEDLINE 92147684
PUBMED 1737784
FEATURES
source

Location/Qualifiers
1..1739
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="W303-1A"
/db_xref="taxon:4932"
/chromosome="XVI"
234..1523
/gene="TFIIfa"
234..1523
/gene="TFIIfa"
/note="zinc finger containing protein"
/codon_start=1
/product="transcription factor IIfa"
/protein_id="AAB0804.1"
/db_xref="GI:172902"
/translation="MGEVLANNEGMPLEIKQETIPISRESESEINSINLSTRESSN
RPEYFCDYDCDKAFTRPSILTEHQLSVQGLFARQCDKASFYKSKLHEHLYTH
SDTKPQCYCGGVTTRQOLKRHEVHTKSFICPEGGCLRYKHQOLAHLLSYHL
HKLTPCHNKSFORPYRLRNHLSKHDPVENPYOCTPAGCKEPRIMQLOSHIKXD
HPKLCRPSCKYGVENGOMMIHDDSLVTNNKCHICPDMSFSRKHLLTHYGS
HREBDEPLEIKRKISDIOQLVDHGVQLGNSKHSNDEKISNLRKRLKENNV
EFLQNVLDLKRLESGENGLNLTNTVGKRYRCFYNNCSRTGTFKRGYKHKHDKHVKH
ELKLTLOREKENKTLVDONHKEPFIIOKETOSAGDK"

ORIGIN

Query Match 10.6%; Score 217.4; DB 8; Length 1739;
Best Local Similarity 59.4%; Pred. No. 4.4e-25;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

QY 781 GTCCAAAAGATATATTTGACATATGAAAGGGTGTGATTAAGCCCTATATGACCATCAT 840
DB 367 GGGCAAGACATATTTCTGTACTATGATGGCTGTGATTAAGGCAATTTACAGACCTTCAA 426
QY 841 TATTAGAGCAACATTTAAGAACCCACAGTATGATGACCGTATTAATGTATACAGTGAAG 900
DB 427 TTTGACTGAACACCAATTAAGGTACATCAGGGTTTA--AGAGCATTTCACTGTGATTA 483
QY 901 ATTGTGATAAGATTTTTCAGAAATCATCTTTGGAACACATATGTATCATATTTCCG 960
DB 484 AGTGTCAAAATCTTCTGTTAAAAAGAGTCACTTGAAGACACATTTGATACGCAATCTG 543
QY 961 AAAAAAACCATTTCATTTGTGATGTGTGTAAGGGGTTAATCTGCAACAACATTTGA 1020
DB 544 ATACGAACCATTTCCATATTTCTTATTTGTGAAAAAGAGTACGACTGCCGACACTGA 603
QY 1021 AAGACATGAATACACCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAG 1080
DB 604 AGCGACACCAAGTAACGATACCAAAATCTTCAATTTGTCCAGAGGAAGATCAACCTCC 663
QY 1081 CATTTTATAACATCAATCTTTAAGACATCATATATATCTGTCATGAAAAACATTTAA 1140
DB 664 GATTTCTACAGCATCCACATTTAAGGGCACATATTTTATCTGTTCAATTTACATTAACATA 1200
QY 1141 CGTGTAACCAATGATTAAGTTTTCACCTGACCTTCAAAATTAAGCAACAATTAATA 1260
DB 724 CCTGTCCACACTGCATAAAGCTTTTCAGAGCCATATAGGCTTAAGAAATTCATTTCTA 783
QY 1201 AACATCA-----TGTGTGATCTCTGTTATCAATGTATCATCTGGTTGTTTAAAA 1254
DB 784 AACATCAAGATCTGTAGGTAGAAAATCTTACCAATGATCTTTGTGTGGTTGTGCAAG 843
QY 1255 ATTTCCAAACCTTGATGATTTACAAATTTCAATTAACAACTGATCCAAACTTAAT 1314
DB 844 AGTTTGGATATGTGATCAATTTGACATGCAATTAATAAATGATCATCTCTAATTAATAAT 903

QY 1315 GTCTTAATGTGGTAAAGTTGTGTGGGAAAAAGTTTATCTTCACATATGTTAAGTC 1374
DB 904 GTCTTATTTTGTAGCAACCATGTGTGGGAAAAAGTTTAAAGCACATATATATTC 963
QY 1375 ATGATGATTTACCATGATCAAAAATATGACTTGTGATTTATGTGTGAGGAAAAATTGG 1434
DB 964 ATGACGACTACAGTACGTAACCAAAAATTGGAAGTGTCTATATATGTCCTGATATATGCTTTT 1023
QY 1435 CAAAGAAAAATGAATTAAGTTGAACATTAT 1463
DB 1024 CTAGAAAAACATGATCTTCTCAGCATTTAT 1052
RESULT 6
YSCP9677/c 37497 bp DNA linear PLN 01-AUG-1997
LOCUS Saccharomyces cerevisiae chromosome XVI cosmid 9677.
DEFINITION U25841 U00094
ACCESSION U25841.1 GI:786295
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted by: Robert Waterston
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: mj@sequencer.wustl.edu
NEIGHBORING COSMID INFORMATION:
This sequence includes nucleotides 1-37497 of cosmid 9677. The
cosmid on the left is 9705.
Location/Qualifiers
1. 37497
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="S288C (AB972)"
/db_xref="taxon:4932"
/map="XVI"
/note="baker's yeast"
complement(2366..3655)
/gene="TPC2"
/gene="TPC2"
/note="p9677.9"
/codon_start=1
/product="Transcription factor TFIID (PIR accession
number S20050)"
/protein_id="AAB64615.1"
/db_xref="GI:786305"
/db_xref="GI:786305"
gene
CDS
SPTKPFQCSYCGKGVTTTRQQLKRRHVVTHTKSFICPEEGCNIRFYKHQPLRAHLLSVHL

HLKTCBCKNSFORPYRLRNHLSEKHDPVENVEYQCTFAGCCKEPIIMSQLOSHIKND
HFKLKPICSPKCVGNGEOMHMIHDDSLVTNMGCHICPDMSFSKRDILTHYGSII
HREBDIPLFKYKISIDISQIYODHGVQDLSGNSKSNODEKTSNRLKRRKILTENNV
EPLQNEVDLEKRLSEBNGNLNLNTVGRKTRCFYNNKCSRTFTKRYEKHIDKHVH
ELKLIKLOEKEENKTLVDQNHKPFIIQKETSAGUK"
3890..4433
/gene="RPB6"
join(3890..3909,3986..4433)
/gene="RPB6"
/note="p9677.8"
/codon_start=1
/product="23 kD polypeptide of RNA Polymerases I, II, and
III (Swiss Prot. accession number P20435)"
/protein_id="AAB64616.1"
/db_xref="GI:786304"
/translation="MSDYEARFNDGNGENEPEDVHEHSPDETYREKPOFGDGTNDAN
GKTIVVGNGPEDFQOHEQTRTKLEKALPKORATTPMTYKERRRIIGTALQIS
MNAFVVDLEGETDPLRIAMKEKELPIVIRYLPDSFEDMSVELIVDL"
complement(4621..5112)
/gene="p9677.10"
complement(4621..5112)
/gene="p9677.10"
/note="Similar to calmodulins (e.g., Swiss Prot. accession
number P13868) and calmodulin-related proteins (e.g.,
Swiss Prot. accession number P27482) from several
organisms."
/codon_start=1
/evidence="not experimental"
/protein_id="AAB64617.1"
/db_xref="GI:786306"
/translation="MDHSESLTFNQDYINKLKAFQMLDEDEDEGLISRGDITKLY
ATGKTLIDDEMSKMPVDESLTAEVEEVSVPFIFSLMGKINSQFPESEBELSESK
ALGRGDHIANVPLNEVIDSLKEAGFENPEEFALKFKFTNQATERTERTRGKLFIDS
ITD"
5297..9595
/gene="SKI3"
5297..9595
/gene="SKI3"
/note="p9677.7"
/codon_start=1
/product="Contains 8 copies of the TPR domain"
/protein_id="AAB64618.1"
/translation="MSDIKOLKEAKQELTNPDYEETISEKVLKLPDNYFAHIFL
GKALSLPASNVSNRNLNATNHYSAKLVADNLNLMKGLPLFRTEVVPDILS
YDERPDCCGYADALIKOBOSYELINDIKLTKTHPDCKQAFQYHOKPSIMMETG
RHLSHPDQALNLNLKILSNIEITRTIGTLQNLKILKASPDYQIKLNSPSEITKNS
EIHQNLQVNLNLTADQKSEITENQWLEYRIKVLKSPLVKQDFTRKMWEDVYL
VNHQSLAMQKYPEMDYEDLDNDADPLIKYFKKPRDPLAMLLYSWLSKSKSYDL
KSLSEANKPEEGHKTKEKENDIKQVDETNEDKQRYDEDEKQVDEVEKQDEAKS
DEEEDDDIEIGLEEIVTVLTENIVKCKNNILAHRLCYLYLTKEVEFALPYIKN
GSLIYANIRIDIGVHPLTRKEPSLDATYATVYDAKDNAAALKYDNTLISGFSNT
QAKKKGKILITERKNMDKMTLLTVYHEGSPNNLEEVLSLMSKAWNGMEATAGLD
TVKIGKMDRSLDFRALNIMROKATYIMKHSINAKOENYCAKLLIOSKILTD
TPAFGFTSLDIYCHYKDHARFCYKPAFDLDAGYTAATKILTEYAFENQOAS
SIASRLIKGEKAKELSNMWPERVAVIHLKQEBDSLEWFSQALVADPNDESVW
GLGQAYACQRIEASIKYFKALQIRSHFAQYFKALSLCDVGEYLESJDLIRKVCQ
EAATESFOIGLVEVLRCSLDLSOGFLKXSIAADTERTKIIISLCEENQOQW
IYLSQVLRPTWESKVDTLPVESIVSIFENSOGSEEDISVNIKILTEPILKEQYDAAI
VNSLAKFLLIASKYSVSDQFTDIAGVYASVYKYNIGISLTPATILKEQYDAAI
FAFKRSIQLOSNITSEWIIGGIATMDINFRVSQCFIKATLRLKRAITNTNMLMGIL
KKQDTEFAQVNLNQLQSLAPQDSFWTGMALILEQGDIIQSSKLPASFLISGRSK
AAGPMAKAVNLNINNGDDEBDIETVEKILITASIAEQFPKSPDSQFALQCALTL
ERLHHYENANLRLIIGLEKKEKODEBELFNPAIKQGFARHILGLENPLSIE
NADLSQGITSSSDEKSMKTKISNHICGLGSYPLNPDQTLNPOELISISKSKIL
VNLIAKVLIVGESDTYKELALQELTETATSGADLVLTITAAISLDDKREDISLIL
EELKALPLSKQIIDLKHDAPYLIEITKRLIRNDTGQVQWRSAYFFPNNLKVRBRLD
KNIQRIASNGQNTVTAEMSKLYCESKSNLSIORGMFLCEWNTAVYALNECP"
complement(9710..11674)
/gene="RPC3"
complement(9710..11674)
/gene="RPC3"
/note="p9677.11"

QY	781	GTCCCAAAAAGTATATTGGCAATATGAAAGGCTGTGATTAACCTTAATATCGACCATCAT	840
Db	3522	GGCCAAAAGCATATATTTCTGTGACTATGATGGCTGTGATTAAGCATTTTACAAAGCCTTCA	3465
QY	841	TATTAGAGCAATTTTAAGAACCCACACAGTAATGATCGACCGGTATAATGTACATGTGACG	900
Db	3462	TTTTGACTGAACCAACATTTAAGCGTACATCAGGGTTTA--AGACATTTTCAAGTGTGATA	3406
QY	901	ATTGTGATTAAGCATTTTTCAGAAAATGACATTTTGGAAACACATATGTATCATCATTCG	960
Db	3405	AGGTGTCAAAATTCCTTCGTTAAAAAGATGCACTTGAGAGACACTGTATATGCAATTCG	3346
QY	961	AAAAAAACATTTCCATTTGTTCACTGTGTGTGTAAGGGGTAAATTTCTGCACAACTTGA	1020
Db	3345	ATACGAAACCATTTCCAAATGTTCTTATTTGTGAAAAAGAGTACGACCTCGCAGCACTGA	3266
QY	1021	AAAGCATGAATTCACCCATTCACAAAGTATTTAAATGTACATTTGAAATTTGTCAAGAG	1080
Db	3285	AGCGACGCAAGATTAAGCATCAATTCATTTTTCAGAGAGAGATGCAACCTCC	3226
QY	1081	CATTWTATAACATCATATCTTTTAAGACATCATATATTATCTGTCATGAAAAACATTTA	1140
Db	3225	GATTCACAGATCCACAAATTTAAGGACATATTTATCTGTTCATTTACATTAACCTAA	3166
QY	1141	CGTGAACCATGTATAATAAGTTTCACTGACCTTCMAAATTAAGCACAAATTAATTA	1200
Db	3165	CCTGTCACACCTGCATTAAGAGCTTTCAGAGGCCATTAAGGCTAAGAAATCACATTTCTA	3106
QY	1201	AACATCA-----TGTGATCTCTCTGTTATGCAATGTGATCATCTCTGTTGTTTAAA	1254
Db	3105	AACATCAGATCTCTGAGGTAAAGAAATCTTTACCAATGTACTTTTGTGTTGTTGCAAG	3046
QY	1255	ATTTCAAACTTGTGTCAATTTTACATTTTCAATTAATAACATGCATCAAACTTAAAT	1314
Db	3045	AGTTTCGATATGTGTACAAATTTGCAATTCGATATTAATAATGATCTCTTAATTTAAAT	2986
QY	1315	GTCTTAATGTGTAAAGTTGTGTGGAAAAAGTTTATCTTCAATATGTTAAGTC	1374
Db	2985	GTCTTAATTTGTAAACCAATGTGTGGGGAAAAAGTTTAAACAAATGACATGATTATTC	2926
QY	1375	ATGATGATTCACATATGAAATAATGACTGTGTATTTGTGATGTGGGAAAAATTTG	1434
Db	2925	ATGACGACTCATGTAAACAAAAATTTGAAAGTCAATATGTCCGATATGTCTTTT	2866
QY	1435	CAAGAAAAATGATTAAGTTGAACATTTAT	1463
Db	2865	CTAGAAAAACATGATCTTCTCAGCGATTAT	2837

Draft entry and computer-readable sequence for [unpublished (1990)]
 kindly submitted
 by J.D. Friesen, 01-MAY-1990.
 The Hospital for Sick Children
 555 University Avenue
 Toronto, Ontario, CANADA M5G 1X8
 e-mail: jfm@sickkids.toronto.edu.
 Location/Qualifiers

1. .2133

/organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"

/db_xref="taxon:4932"

join(1026..1045,1122..1569)

/note="RNA polymerase II sixth subunit (RP026)"

/codon_start=1

/protein_id="AAA34989.1"

/db_xref="GI:1172453"

/translation="MSDYEEAFNDGNEFDVHEHSDSEETEEKPFQKDETDAN
 GRTVTGNGPEDFQOEIRKTKLEKAIIPDQRTPTMYTKYERARILGTRALQIS
 NNAFVFDLEGETDPLRIAMKELEKTIPLVIRRIIDPSFEDMSVEELIVDL"

<1026..1045

/note="RNA polymerase II sixth subunit (RP026)"

/number=1

1046..1121

/note="RP026 intron A"

1122..>1569

/note="RNA polymerase II sixth subunit (RP026)"

/number=2

ORIGIN

Chromosome XVI.

Query Match 9.7%; Score 199.2; DB 8; Length 2133;

Best Local Similarity 61.4%; Pred. No. 3.3e-22;

Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

890 TAGACGACGATGTCATTAAGCATTTTCAGAAATCAATTTGGAAACATATTTGT 949
 556 TTCAAGTGTAGTAGTGCAAAATCTTCGTTAAAGACACTTAGAGAGACTTGTGA 497
 950 ATGACATTCGAAAAAAACCATTCCATTGTTCACTGTGTGTGTGAAGGGTTAATTTCG 1009
 496 TAGGCATTTCGATCGAAACCATTCATGTTCTTATTGTGAAAAAGAGTACGACTCG 437
 1010 ACAACACTTGAAGACATGAATACCCCTACACAACTCAATTTAAATGTCATTTGAAA 1069
 436 CAGACCACTGAAGGACAGAGTACGACATACCAAACTTTCATTTGTCCAGGAAAG 377
 1070 TTGTCAAGAGCATTTTAAATCAATCTTTAAGACATCATATTTATCTGTTCAATGA 1129
 376 ATGCAACCTCGATTCACAAAGCATTCACATTTAAGGACATATTTATCTGTTCAATT 317
 1130 AAAACATTTAAGCTGAACATGTATTAAGTTTCACTGACCTTCAAAATTAGACA 1189
 316 ACATTAACCTACCTGTCACACTGCATTAAGAGCTTCAGAGCCATTAAGCTTAAGAAA 257
 1190 ACATTAATTAAGATCA-----TGTGATCTCCGCTTATCATGATGATCACTCTCG 1243
 256 TCACATTTCTAAACATCGATCTCGAGGTGAAAAATCTTATCAATGATGATCTCTCG 197
 1244 TTGTTTAAAAATTTCCAACTGGTCAGTATTCAATTTCAATATTAACCAACATGATCC 1303
 196 TTGTTGCAAGAGTTTCGATATGTCACAAATGCAATCGATATGAATTAAGATCATCC 137
 1304 AAAACCTTAATGTCTTAATGTGTGAAGTTGTGTGTGGAAAAAGTTTATCTTCACA 1363
 136 TAAATTAATTAATGCTCTATTGTAGCAACCAATGTGTGGGAAAAATGTGTTTACAAATGCA 77
 1364 TATGTTAAGTCAATGATTTCAACATGATCAAAATATGACCTTGTGATTTATGATGAT 1423
 76 CATGATTTATTAAGACACTCACTAGTAAACAAAAATTTGAAGTGTCAATATATGTCCTGA 17
 1424 GGGGAAATTTGCAA 1437
 16 TATGCTTTTCTTA 3

RESULT 8

CNS061NG

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

CNS061NG 797 bp DNA linear STS 30-NOV-2001
 T7 end of clone AS0A022B07 of library AS0AA from strain CLIB 533
 of Saccharomyces bayanus, sequence tagged site.

AL400514

AL400514.1 GI:12156635

ST8.

Saccharomyces bayanus

Saccharomyces bayanus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 797)

Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,

Boillot-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,

de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,

Malperry,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,

Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

Wincker,P. and Weissenbach,J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

2 (bases 1 to 797)

Bon,E., Neuvéglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,

Aigle,M. and Durrens,P.

Genomic exploration of the hemiascomycetous yeasts: 5.

Saccharomyces bayanus var. uvarum

FEBS Lett. 487 (1), 37-41 (2000)

3 (bases 1 to 797)

Genoscope.

Direct Submission

Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This STS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

Location/Qualifiers

1..797

/organism="Saccharomyces bayanus"

/mol_type="genomic DNA"

/strain="CLIB 533"

/variety="uvarum"

/db_xref="taxon:4931"

/clone="AS0A022B07"

/clone_lib="AS0AA"

/note="Tend : T7"

<2..>796

/note="similar to Saccharomyces cerevisiae ORF YPR186C [

PZFI : YPR186C (transcription initiation factor)]"

/evidence=not_experimental

misc_feature

ORIGIN

Query Match 9.4%; Score 193; DB 11; Length 797;

Best Local Similarity 59.3%; Pred. No. 3.7e-21;

Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;

897

GACGATTTGTAAGCATTTTTCGAAATCAATTTGGAACAACATTTGTATCAT 956

83

GACCAATTGTGCAAAATCTTTGTTAABAAGTCACTAGAGAGGCACTTTTCGAT 142

Query 957 TCCGAAAAAACCATTCATTTGTCAGTGTGTTAAAGGGTTAATTCGCAACAC 1016
Db 143 TCTATTCGAGAGCCATTCATTTGTTATTTGTCAGAGGGGGTGAACAACCTGCACAGCAA 202
Qy 1017 TTGAAAAGACATGAAATCACCACATCAAAAGTCATTAAATGTAATTTGAAATTCGCA 1076
Db 203 TTGAAGGCGCCAGAAAGTTACACATACCAAACTTTTGTTCGGAAGATGATGAT 262
Qy 1077 GAACATTTTAAACATCAATCTTTAAAGACATATATTTATCTGTTCAAAAAACA 1136
Db 263 CTCGATTTTAAAGACCCCTCAATTAAGGGCTCATTTCTGTTCACTGACCTAGACAA 322
Qy 1137 TTAAGTGTAAACATGTAATTAAGTTTCACTGACCTCAAAATTAAGACACATAAA 1196
Db 323 TTAACCTGTCCACATGATGATTAATTTTCAAGAGACCTCAAGGGCTTAAGAAATCATATC 382
Qy 1197 TTAACATCATGCT-----GGATCTCTGCTTATCATGATGATCATCTGCTGTTT 1250
Db 383 TCTAAACATCATGATCTGTAAGGTGAATCCGATTAATGATCTTCTGCTGCTGTTGT 442
Qy 1251 AAAAATTTCCAACTGTGTGATGATTAACAATTTCAATTAACAACTGCATCCAAAATT 1310
Db 443 ACAGAGTACCGTATATGTCACAGTTGCATCATATCAAAATGACATCTTAATTTG 502
Qy 1311 AATGTCTTAATAGTGTAAAGGTGTGTTGGAAAAAGTTTATCTTCAATATGTTA 1370
Db 503 AATGCCCCATTTGCAACAAGCCTGTGTGAGGGGAAATGTCTCAAAATGATGATGATC 562
Qy 1371 AGTCATGATGATCTACATGATCAAAATATGACATGATGATGATGATGATGATGATGAT 1430
Db 563 ATTCATGATGATCTTGTGTAATGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 622
Qy 1431 TTGCAAAAGAAAAATTAATTAAGTTGATGACATTAATTAATGTTCTCAT 1475
Db 623 TTCTTGAAGAACGACGACCTCTTACTCATTACACAGATGACGAT 667

RESULT 9
AC138126/c 169063 bp DNA linear PRI 17-DEC-2002
LOCUS AC138126
DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
AC138126
VERSION AC138126.1 GI:27151357
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 169063)
DOI Direct Submission
DOI Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS Unpublished
TITLE 2 (bases 1 to 169063)
JOURNAL Direct Submission
DOI Direct Submission
DOI Joint Genome Institute and Stanford Human Genome Center.
COMMENT Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov

FEATURES
source
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

Location/Qualifiers
1. 169063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-274A19"

ORIGIN

Query Match 4.9%; Score 101; DB 9; Length 169063;
Best Local Similarity 50.8%; Pred. No 6.1e-07;
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

Qy 726 GAAAGTACGAAACCAATGATATCATCTTTATATCTTCTTCTTCATCAGTCCC 785
Db 80134 GAAATGGGCAAACTTTAAACAGTCTCATCTTACTACTAGAAATTCATACT 80075
Qy 786 AAAAGATATTTGACATATGAAAGGTGATTAAGCCATTAATGACATCTTA 845
Db 80074 GGAAGAAATCTTCAAAATGAAATGGAATGGGAAAGCTTTATATCATCTTCAAACTT 80015
Qy 846 GAGCAACATTTAAGAACCCAGATTAATGATGACAGCCGATTAATGATGACAGTATGT 905
Db 80014 ACTGAACATTAAGAAATTCATCTGAGAGAAACCTTACATGAT-----GAAGATGT 79961
Qy 906 GATTAAGCATTTTTCAGAAAATGACATTTGGAACACATATGATCATCATTCGAAAAA 965
Db 79960 GGCAAAGCTTTTAAACATTTCTCAGATCTTCTGACATAGATTAATCTACTGAGAG 79901
Qy 966 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTGACACACCTTGAAGA 1025
Db 79900 AAACCTTACAAATGTAAGAAATGTGTAAGGCTTTAAACAGTCTTCAACCTTACTAG 79841
Qy 1026 CATGAATCACCACATCAAAAGTCATTAATGATGATTTGAATAATTTGCAAGCATTT 1085
Db 79840 CATTAAGAAATTCATCTGAGAGAAACCTTACCAATGTAAGAAATGTGCAAAAGCTTTT 79781
Qy 1086 TATTAACATCAATCTTTAA--GACATCATATATATCTGTTCAAGAAAAACATTAAG 1142
Db 79780 AACCAAGTCTCAAACTTACTGACATTAAGAAATTCATCTGTTGAGAAACCTTCAAA 79721
Qy 1143 TGTAAACATGTAATTAAGTTTTCAGTCTTCAAAATTAAGACACATTAATTA 1202
Db 79720 CTTAAAGATTTGAAGATGATTTGAAACACCTTCAAAAGTTTCTTAACATTAAGAAAT 79661
Qy 1203 CATCATGTGATCTCTGCTTATCATGATGATCATCCGTTGTTTAAATTTCCAA 1262
Db 79660 TATGCTGTGAGAAATCTTGAAGATGTAAGATGTAACAAACCTTTAAAGGTGTGAC 79601
Qy 1263 ACTTGCTGATTAACAATTTCAATTAACAACTGCATCCAAACT 1309
Db 79600 ACTTGATTTGCTATGAATTAATTTCACTTAAGAAAAAACCTTCAAGT 79554

RESULT 10
AC073544/c 177299 bp DNA linear PRI 19-JUN-2002
LOCUS AC073544
DEFINITION Homo sapiens chromosome 19 clone RP11-359H18, complete sequence.
AC073544
VERSION AC073544.4 GI:21465367
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 177299)
DOI Direct Submission
DOI Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS Unpublished
TITLE 2 (bases 1 to 177299)
JOURNAL Direct Submission
DOI Direct Submission
DOI Joint Genome Institute.

COMMENT Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 19, 2002 this sequence version replaced gi:13699752.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov

REFERENCE AUTHORS	
1	
Ninomiya, K.,	Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H.,	Furuya, T., Takahashi, M., Kikawa, E., Omura, Y., Abe, K.

Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suuki, Y., Sugano, S., Nishihara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2132)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO Human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-& 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

Location/Qualifiers
1..2132
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2R12000341"
/cell_line="NT2"
/cell_type="keratocarcinoma"
/clone_lib="NT2R12"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction..majorly NT2 neuron"
148..1878
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71090.1"
/db_xref="GI:16551398"
/translation="MVAPPVICSHPAHPSPFSPONIKDSFOKVTTPRYKCEHENTL
SKSVDECKYQKNGYNGLCPLTPQSKIFQCDKYMKI PFKFSNLNGHKYRHRKRPK
YKEFGSCFIPSNLTQHKICTRVNPKYKEDGKAFNGSSITTKRKHITGKSTICE
EGKANOFNTLITTKITITRDKLYKRECSAFNLSHITHTTIIHTGENPKYKEEC
DKAFNOSSTLTTHKTIHTREKLINEKYEKCGKAFNOSHTLTKRKHITGKPYCEEGK
AFNOSHTLTKRKHITGKPYCEEGKAFNOSHTLTKRKHITGKPYCEEGKAFN
NSHSLTKRKHITGKPYCEEGKAFNOSHTLTKRKHITGKPYCEEGKAFN
FNSLTKRKHITGKPYCEEGKAFNOSHTLTKRKHITGKPYCEEGKAFNOSHTL
KLTEHKIHTGKPYCEEGKAFNOSHTLTKRKHITGKPYCEEGKAFNOSHTL
TKRKHITGKPYCEEGKAFNOSHTLTKRKHITGKPYCEEGKAFNOSHTL
HKRNYAGEKS"

CDS

Query Match 4.8%; Score 99.4; DB 9; Length 2132;
Best Local Similarity 50.6%; Pred. No. 2.3e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
Db 726 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTCTCTTCATCAGTCC 785
1384 GAATGTGGCAAGCTTTTAACAGTCCCAATCTTACACACATTAAGATTTACTACT 1443
Qy 786 AAAAGTATTTTGCATATGAAAGGTGTGATAAAGCTTATATGACCATCTATTATTA 845
Db 1444 GAGAGGAAATCTTACAAATGTGTAAGATGTGCAAAAGCTTCTATGATCTCCAAACTT 1503
Qy 846 GAGCAACATTTAAGAACCCACAGTATGATGACGATTAATATGACAGTGAACATTTGT 905
Db 1504 ACTGAACATTAAGAAATCTTACTGAGAGAAACCTCTACATGT-----GAAAGATGT 1557
Qy 906 GATAAGCATTTTTCAGAAATTCACATTTGAAACACATATTTGATTCACATTTCCGAAAA 965
Db 1558 GCGAAAGCTTTTAAACACTCTCTCAACCTTGCTACACATTAAGTATTTCACTGAGAG 1617

ORIGIN

Query Match 4.8%; Score 99.4; DB 9; Length 2132;
Best Local Similarity 50.6%; Pred. No. 2.3e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
Db 726 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTCTCTTCATCAGTCC 785
1384 GAATGTGGCAAGCTTTTAACAGTCCCAATCTTACACACATTAAGATTTACTACT 1443
Qy 786 AAAAGTATTTTGCATATGAAAGGTGTGATAAAGCTTATATGACCATCTATTATTA 845
Db 1444 GAGAGGAAATCTTACAAATGTGTAAGATGTGCAAAAGCTTCTATGATCTCCAAACTT 1503
Qy 846 GAGCAACATTTAAGAACCCACAGTATGATGACGATTAATATGACAGTGAACATTTGT 905
Db 1504 ACTGAACATTAAGAAATCTTACTGAGAGAAACCTCTACATGT-----GAAAGATGT 1557
Qy 906 GATAAGCATTTTTCAGAAATTCACATTTGAAACACATATTTGATTCACATTTCCGAAAA 965
Db 1558 GCGAAAGCTTTTAAACACTCTCTCAACCTTGCTACACATTAAGTATTTCACTGAGAG 1617

Qy 966 AACCATTCATTCGTAGTGTGTGTAAGAGGGGTATTCGACCACTGTGAAGA 1025
Db 1618 AACCATTCATTCGTAGTGTGTGTAAGAGGGGTATTCGACCACTGTGAAGA 1677
Qy 1026 CATTAATTCACCATTAACATGATTTTAAATGTACATTTGAAAAATGTCAGAAACATTT 1085
Db 1678 CATTAATTCACCATTAACATGATTTTAAATGTACATTTGAAAAATGTCAGAAACATTT 1737
Qy 1086 TATTAACATCAATCTTTA---GACATCATATATATCTGTTCATGAAAAACATTAACG 1142
Db 1738 AACCATTCATTCGTAGTGTGTGTAAGAGGGGTATTCGACCACTGTGAAGA 1797
Qy 1143 TGTAAACATGATTAAGTATTTTACCTGACCTTCACCAATTTAGACAACTTAATTTAAA 1202
Db 1798 CTTAAAGATGTACATGATTTTGAAGAAACATTTTCTTAACATTAAGAAAT 1857
Qy 1203 CATCATGTGATCTCTGCTTTCATATGATCATCTGATTTTAAAAATTTCCAA 1262
Db 1858 TATGCTGTGAGAAATCTTGAAGATGTAAGAAATGTAACAAACCTTTAAAGTTGTAC 1917
Qy 1263 ACTTGACATATTAACATTTTCAATTAACCACTGATCCAAACT 1309
Db 1918 ACTTGATGTGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1964

RESULT 13
AC099500/c 138627 bp DNA linear PRI 29-MAY-2002
LOCUS Homo sapiens chromosome 19 clone RP11-20906, complete sequence.
DEFINITION AC099500
VERSION AC099500.2 GI:21240690
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 138627)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
2 (bases 1 to 138627)
DOE Joint Genome Institute.
TITLE Direct Submission
Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL On May 29, 2002 this sequence version replaced gi:16930916.
Draft Sequence Produced by DOE Joint Genome Institute

COMMENT

www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.1.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 153.8kb). It is clipped at the overlap with AC008981.
The number of bases overlapped is 27192.

FEATURES

Location/Qualifiers
1..138627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-20906"

ORIGIN

Query Match 4.8%; Score 99.4; DB 9; Length 138627;
Best Local Similarity 50.6%; Pred. No. 1.1e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;


```

QY 726 GAAAGTACGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCAGCTGCC 785
DB 48699 GAATGTGGCAAAAGCTTTTAAACGAGTCTCAATCTTATCAACATTAAGAAATTCATCT 48640
QY 786 AAAAATATATTTTGCACATATGAAAGGGTGTGATTAAGCCCTATATGACCATCATTTA 845
DB 48639 GAGAGAAATCTCAACAAATGTGAAGATGTGGCAAAAGCTTCTATGATCTCAAAACTT 48580
QY 846 GAGCAACATTTTAAGAACCCAGATTAATGATGACCGATTAATATGTAAGTGAAGCATTTGT 905
DB 48579 ACTGAACTATAGAAATTCATTAATCTGAGAGAAACCTACACATGT-----GAAGATGT 48526
QY 906 GATTAAGCATTTTTCAGAAAAATGACATTTGGAACACATATGTAATGACATTCGAAAA 965
DB 48525 GGCAGAGCCCTTAAACATTCCTCAACCTTGCTACATGATGATTAATTCATCTGAGAG 48466
QY 966 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACACATTTGAAAAAG 1025
DB 48465 AAACCTACCAATGTGAAGATGTGTGAAGCCCTTAAACAGTCTCAACCTTACATGA 48406
QY 1026 CATGAATATCCCATTAACCAAGTCATTTAAATGTAACATTTGAAAAATGTCAAGACATTT 1085
DB 48405 CATAGAGATTCATTAATCTGAGAGAAACCTACCAATGTGAAAAATGTGCAAAAGCTTTT 48346
QY 1086 TATTAACATCAATCTTTAA---GACATCATATATTAATCTGTCATGAAAAAACATTAACG 1142
DB 48345 AACCATGCTCAACCTTACTGACATTAAGAAATTTATCTGTGTGAAGAACTCTACAA 48286
QY 1143 TGTAAACATGTATTAATTAAGTTTTCATCTGACCTTCANAAATTTAGCAACATTAATTA 1202
DB 48285 CCTAAAGATGTAAACATGTATTTGAAAAACCTTCANAAAGTTTCTAAACATTAAGAAAT 48226
QY 1203 CATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAATTTCCAA 1262
DB 48225 TATGCTGTGTGAATCTTGAATGTGAAGATGTGAACAAACCTTTAAAGTTGTAC 48166
QY 1263 ACTGTGTCAGTATTAATTAATTCATTAATTAACATGATGATCAAAACT 1309
DB 48165 ACTGATTTGTGATTAATTAATTCATTAATTAACCTTAATAGT 48119

RESULT 14
AC138469/c 169500 bp DNA linear HTG 08-JAN-2003
LOCUS Homo sapiens chromosome 19 clone RP11-189C24, *** SEQUENCING IN
AC138469
AC138469.1 GI:27544967
AC138469
HTG: HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 169500)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 169500)
DOE Joint Genome Institute and Stanford Human Genome Center.
Submitted (08-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
* NOTE: This is a "working draft" sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 75796: contig of 75796 bp in length
* 75797 75896: gap of unknown length

```

```

* 75897 138071: contig of 62175 bp in length
* 138072 138171: gap of unknown length
* 138172 157899: contig of 19728 bp in length
* 157900 157999: gap of unknown length
* 158000 166058: contig of 8059 bp in length
* 166059 166158: gap of unknown length
* 166159 169500: contig of 3342 bp in length.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 479013
Center clone name: RP11-189C24
-----
Summary Statistics
Consensus quality: 165586 bases at least Q40
Consensus quality: 165891 bases at least Q20
Consensus quality: 166156 bases at least Q20
Estimated insert size: 169095000; agarose-fp estimation
Estimated insert size: 166687; sum-of-contigs

estimation
Quality coverage: 0 in Q20 bases; agarose-fp
Quality coverage: 1 in Q20 bases; sum-of-contigs
estimation.
location/Qualifiers
source 1.169500
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-189C24"

ORIGIN
Query Match 4.8%; Score 99.4; DB 2; Length 169500;
Best Local Similarity 50.6%; Pred. No. 1.1e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 726 GAAAGTACGAAACCAATGATATCATCTTTAATATCTTCTTCATCAGCTGCC 785
DB 49667 GAATGTGGCAAAAGCTTTTAAACGAGTCTCAATCTTATCAACATTAAGAAATTCATCT 49608
QY 786 AAAAATATATTTTGCACATATGAAAGGGTGTGATTAAGCCCTATATGACCATCATTTA 845
DB 49607 GAGAGAAATCTCAACAAATGTGAAGATGTGGCAAAAGCTTCTATGATCTCAAAACTT 49548
QY 846 GAGCAACATTTTAAGAACCCAGATTAATGATGACCGTATTAATGTAAGTGAAGCATTTGT 905
DB 49547 ACTGAACTATAGAAATTCATTAATCTGAGAGAAACCTACACATGT-----GAAGATGT 49494
QY 906 GATTAAGCATTTTTCAGAAAAATGACATTTGGAACACATATGTAATGACATTCGAAAA 965
DB 49493 GGCAGAGCCCTTAAACATTCCTCAACCTTGCTACATTAAGTATTAATTCATTAAGTGAAG 49434
QY 966 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGACAAACATTTGAAAAAG 1025
DB 49433 AAACCTACCAATGTGAAGATGTGTGAAGCCCTTAAACAGTCTCAACCTTACATGA 49374
QY 1026 CATGAATATCCCATTAACCAAGTCATTTAAATGTAACATTTGAAAAATGTCAAGACATTT 1085
DB 49373 CATTAAGAAATTCATTAATCTGAGAGAAACCTACCAATGTGAAAAATGTGCAAAAGCTTTT 49314
QY 1086 TATTAACATCAATCTTTAA---GACATCATATATTAATCTGTCATGAAAAAACATTAACG 1142
DB 49313 AACCATGCTCAACCTTACTGACATTAAGAAATTTATCTATCTGAGAGAACTCTACAA 49254
QY 1143 TGTAAACATGTATTAATTAAGTTTTCATCTGACCTTCANAAATTTAGCAACATTAATTA 1202
DB 49253 CCTAAAGATGTGAAGATGTATTTGAAAAACCTTCANAAAGTTTCTAAACATTAAGAAAT 49194
QY 1203 CATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAAAATTTCCAA 1262

```

```

Db      49193  TATCTGCTGAGAAATCTTGAAGATGTGAAGATGTACAAACCTTTAAAGTTGTAC 49134
QY      1263  ACTTGCGACGATTCATTCATTAATAACACTGCATCCAAACT 1309
Db      49193  ACTTGATGTGCAATAGATTAATTCATCTAAAAAAACCTATAGT 49087

RESULT 15
AK122869
LOCUS   AK122869
DEFINITION Homo sapiens cDNA FLJ16502 fis, clone FEBR200664, moderately similar to Zinc finger protein 43.
ACCESSION AK122869
VERSION   AK122869.1 GI:34528067
KEYWORDS  oligo capping, fis (full insert sequence).
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS   Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saio, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Nagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
TITLE     NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 2237)
AUTHORS   Isogai, T. and Yamamoto, J.
TITLE     Direct Submission
JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES
     source          1..2237
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="FEBR200664"
                     /tissue_type="brain"
                     /clone_id="FEBR2"
                     /dev_stage="fetal"
                     /note="cloning vector: pME18SFL3"

ORIGIN
Query Match      4.8%; Score 98.2; DB 9; Length 2237;
Best Local Similarity 52.0%; Pred. No. 3.6e-06;
Matches 273; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

QY      788  AAAGTAAATTTGCACATATGAGAGGTGTGATAAAGCTATATATGACCATATTATTAGA 847
Db      1652  AGAGAAATCTTACAAATGTGAAGATGTGGCAAGCTTTCTATGATCCTCAAAAATTAC 1711
QY      848  GCAACATTTAAGAACCCACAGTATGATCGACCGTATTAATGTACAGTGAAGCGATTGTGA 907
Db      1712  TGAACATAAGAAATTTATCTGAGAGAAACCTTACACATGT-----GAAAGATGTGG 1765
QY      908  TAAAGCAATTTTCAAGAAATCAATTTGGAACAACATATTGTATCAATTCGCAAAAAA 967
Db      1766  CAAAGCCTTAACCATCTCACACCTTGTCTAGACATAGATAATTCACTGTGAGAGAA 1825

```

```

QY      968  ACCATTCATTTGTCAGTGTGTGTAAGGGGTTAATTCGCAACACTTGAAGACA 1027
Db      1826  ACCCTACATATGGAAGAAATGTGTGTAAGGCTTTAACAGTCTCTACACCTTCTACAGA 1885
QY      1028  TGAATTCACCCATACAAAGTCAATTTAAATGTACATTTGAAAATTTGCAAGAGCATTTTA 1087
Db      1886  TAAAGAAATTCATATCTGAGAGAAACCCATCCATGTGAAAATGTGGCAAGCTTTAA 1945
QY      1088  TAAACATCAATCTTTAA---GACATCATATATATCTGTTCATGAAAAAAACATTAAGTG 1144
Db      1946  CGATCTCTCAAAACCTTACTGACATAGAAATTTCTAGTGAGAAACCTTACAAACC 2005
QY      1145  TAAACATGTAAATTAAGTTTTCATCTGACCTTAAATAATAGCAACATTAATTAACA 1204
Db      2006  TAAAGATGTAAAGATTTGAAAACCTTCAAAAGTTTCTTAAACATTAAGAAATTA 2065
QY      1205  TCATGTGTGATCTCTGCTTATCAATGTGATCATCTCTGTTGTTTAAAAATTTCCAAAC 1264
Db      2066  TGCTGTGTGAAGAAATCTTGAAGAAATGTGAAGAAATGTACAAAAACCTTTAAAGTTGTACAC 2125
QY      1265  TTGTCAGATTTACATTTTCATATTAACAACTGCATCCAAACT 1309
Db      2126  TTGATTTGCATTAAGATAATTCATCTATAAAAAAAACCTATAGT 2170

```

Search completed: July 28, 2004, 17:19:28
Job time : 8124 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 28, 2004, 14:00:52 ; Search time 5411 Seconds

(without alignments)
11368.710 Million cell updates/sec

Title: US-09-831-804-1
Perfect score: 2060

Sequence: 1 cttatattaggaagatggc.....gacatactactcttcaatg 2060

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

EST: *
1: em_estda: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estcom: *
17: em_gss_hum: *
18: em_gss_inv: *
19: em_gss_pln: *
20: em_gss_vrt: *
21: em_gss_fun: *
22: em_gss_mam: *
23: em_gss_mus: *
24: em_gss_pro: *
25: em_gss_rod: *
26: em_gss_phg: *
27: em_gss_vrl: *
28: gb_gss1: *
29: gb_gss2: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
C 1	182.2	8.8	628 28	BZ298000 CG3957.f1
C 2	167.6	8.1	959 29	AL405671 T7 end of
C 3	98.2	4.8	1609 11	AK032220 Mus muscu
4	97.8	4.7	2647 11	BC047646 Homo sapi

Result No.	Score	Match Length	DB ID	Description
C 5	97.6	4.7	745 9	AU123448
C 6	92.4	4.5	1101 29	CNS0039G
7	92.4	4.5	2672 11	BC036394
8	92	4.5	591 28	AQ005136
9	91	4.4	593 28	B99387
10	91	4.4	1163 13	BX414627
11	90.4	4.4	2253 11	BC028252
12	90.4	4.4	2694 11	BC022527
13	90.4	4.4	2698 10	BC037782
14	89.2	4.3	604 10	BE161630
15	89.2	4.3	3609 11	BC032590
16	88.6	4.3	535 9	AL705393
17	88.6	4.3	2174 11	AK033001
18	88.6	4.3	2330 11	BC037426
19	88.6	4.3	2622 11	AK033958
20	88.4	4.3	1084 13	BX456765
21	88	4.3	895 13	BUI95416
22	87.6	4.3	832 13	BX437291
23	87.6	4.3	1201 13	BX355654
24	87.4	4.2	1200 13	CG757503
25	87.4	4.2	1392 29	CG757503
26	87.2	4.2	579 28	AQ347265
27	87.2	4.2	687 28	AQ389266
28	87.2	4.2	907 13	BQ423752
29	87	4.2	660 14	CD770140
30	87	4.2	842 13	BUS07408
31	86.8	4.2	521 28	AQ475626
32	86.8	4.2	705 28	AQ194282
33	86.8	4.2	860 28	AQ749175
34	86.8	4.2	897 13	BQ431141
35	86.8	4.2	1906 11	AK013043
36	86.6	4.2	694 12	BG696505
37	86.6	4.2	1098 13	BX377526
38	86.4	4.2	874 13	BUI771853
39	86.4	4.2	2791 11	BC020045
C 40	86.2	4.2	523 14	CA874049
C 41	86	4.2	557 10	BR817669
C 42	86	4.2	667 14	CP135920
C 43	86	4.2	685 29	AG149704
44	86	4.2	801 14	CD657347
C 45	85.8	4.2	1201 13	BX458623

ALIGNMENTS

RESULT 1
LOCUS BZ298000/c 628 bp DNA linear GSS 31-OCT-2002
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata
ACCESSION BZ298000
KEYWORDS genomic clone CG3957, genomic survey sequence.
SOURCE BZ298000.1 GI:24440936
ORGANISM GSS:
Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycetia; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), R10 (2003)
MEDLINE 22508158
PUBMED 12620120
COMMENT
Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.

FEATURES

Location/Qualifiers
1. 628

/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG3957"
/clone_1ib="Candida glabrata Random Genomic Library"

ORIGIN

Query Match 8.8%; Score 182.2; DB 28; Length 628;
Best Local Similarity 59.5%; Pred. No. 4.7e-18;
Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

QY ATATGAAGGCTGATTAAGCCCTTAATGACACCATTAATTAGACAACTTAAAGAC 862
DB ATACGATTAATTGACAGAGCCCTTTCAGAGCCGCTCTCTACAGAAACACAGAAATAC 566
QY 863 CCAAGATGATGACAGCCGTAATAAGTACAGTGCAGATTTGATTAAGCATTTTTCAG 922
DB 565 ---AGTGCATCTTGGGAGAAACCTTGAATGTAAATGTAAAGTTCACTTAA 509
QY 923 AATATCACTTTGGAACACATATTTGATGACATTCGAAACAAACCATTCATTGTTTC 982
DB 508 AAGGATCCACTTGAAGAGACCTTATACACACAGATGAAGACCGTTTATTTTC 449
QY 983 AGTGTGTGAAGGGGTAAATTCGACAAACCTGAAAGACATGAATACCCATAC 1042
DB 448 ATTTTGGGAGAGGGCTAAATTACTAGCAACACTGAACATGAGTAACTCACAC 389
QY 1043 AAGATCACTTAATGTATCTTGAATTTGTCAGAAAGCATTTTATTAACATCAATCTT 1102
DB 388 CAATCGCTTCAATGTGAATATGAAGGTGTGAAGAGTTTCAAGACCCCTCAAT 329
QY 1103 AAGACATCATATATTTATCTGTTTCAGAAACAACTAAGTGAACATGTATAAAGT 1162
DB 328 AAGGGCTCAATTTTACAGATCTTTCACAAAGCTTAATATGACGATGAACAAAG 269
QY 1163 TTTCATCGACCTTCAAAATTTAGCAACATTAATTAACATATGATGATCTCC-- 1219
DB 268 TTTTCAAGACCTTACAGACTCAAAATTCACATGACCAACACATATCCAGATGTGT 209
QY 1220 ---TGTATTAATGATGATCATCTGCTGTTTAAATTTCCAACTTGTCAGATTT 1276
DB 208 CAATGCTTATCAATGTCTTCAAGTGTGTTTCAAGGTTCAAAATATGATGTCAGT 149
QY 1277 ACAATTCATATTAACACATGACATCCAAACTTAATATGTCCTTAATGTGTAAGGTTG 1336
DB 148 AAGATTACATGTTAAATATGATCATCGAAGTTAAATATGCGCTTATATGCAATTAACCTG 89
QY 1337 TGTGGGAAAAAAGTTTATCTTCAATATGTTAAAGTCATATGA 1381
DB 88 TGTGGGGAAGACGGTTTAAATATGACATGAATAATCCACATGA 44

RESULT 2
CNS06MMP/c 959 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AU0AA005F10 of library AU0AA from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION ALA05671.1 GI:12168715
VERSION ALA05671.1 GI:12168715
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
1 (bases 1 to 959)
Souchier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bojotlin-Pukhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,U., Dujon,B., Durieux,P., Lepingle,A., Llorente,B.,
Malperuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Sautin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

TITLE

Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

Neuvéglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Gallardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
FEBS Lett. 487 (1), 56-60 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES

source

1. 959
/organism="Saccharomyces kluyveri"
/mol_type="genomic DNA"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0AA005F10"
/clone_1ib="AU0AA"
/note="end : T7"
complement (<71..>958)
/note="similar to Saccharomyces cerevisiae ORF YPR186c [PZ1 ; TFI1A (transcription initiation factor)]"
/evidence=not_experimental

ORIGIN

Query Match 8.1%; Score 167.6; DB 29; Length 959;
Best Local Similarity 57.2%; Pred. No. 5.7e-16;
Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;

QY 930 CATTGGAACACATATTTGATATCAATTCGAAAAAACCATTCCATTGTCAGTGT 989
DB 958 CACTTGGAAGACATATGTTACGATTCGATTAATAACATTTCAATGTGTCATGTGT 899
QY 990 GGTAAAGGGGTTAATTTCTGCAACAACCTTGAAGAAACATGAATACCATTAACAAGTCA 1049
DB 898 GGTAAAGGTTTACTACCAAGACAGATTAAGAAACATGAGATCAACATCAATATG 839
QY 1050 TTTAATGTATCACTTGAATATTTGTCAGACATTTTAATCAATCAATCTTTAAGACAT 1109
DB 838 TTCAATGTATATGAAGGCTGCAAGCAAGTCTTTCACAGATTCCTCAATAGTCA 779
QY 1110 CATATATTATCTGTTTCATGAAGAAACATTAAGTGAACATGATTAATTAAGTTTCACT 1169
DB 778 CATACACTTGGTATCACTTGCAAGAAATGACCTGTGAGCATTTGGGTAATAATGTTCA 719
QY 1170 GCACCTTAAATTTAGCACACATTAATTAAGCAATCATGATGATCTCCCTGCT----- 1223
DB 718 AGGCATATATCTGTCAGAAACCATCTTCAAGATCAACATGATGATGATTAAGCA 659
QY 1224 TATCAATGATATCAATCCGTTGTTTAAATTTCCAACTTGCTAGATTAATTAATTT 1283
DB 658 TATCAATGATTAACAGGCGGTGACACAGACATTTTAATTAACCTGACATGATGATGATG 599

QY 1284 CATATAAAGCACTGCATCCAAACTTAAATGCTTAAGTGTGTAAGTGTGGG 1343
 Db 598 CATTCAGGCGGACCAATCCCAAGTTGACGTGAACGTATGTGGCAAGCTGTGGT 539
 QY 1344 AAAAAGGTTATCTTCATATGTAATGATGATGATGATGATGATGATGATGATG 1403
 Db 538 GAGCTGTGTCTACGAGATGACATCGAGTCCATACACAGGCTTTAGTAATTAAGTGG 479
 QY 1404 ACTTGATATTTGTGTGATGTGGGAAATTTGCAAGAAATGAATTAAGTGAATTAAT 1463
 Db 478 AAATGTACCAATTTGTGACGAAGTCTTTTGTCTAAGAAAGGTGATCTTTATCTCACTAC 419
 QY 1464 AATATCTTCATATGATGATTAATGCC 1489
 Db 418 ATGGAACATCATTAAGATGATATACC 393

RESULT 3
 AK032220
 LOCUS
 DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
 accession AK032220.1 GI:26328050
 VERSION AK032220
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komuro, H., Akiyama, J., Nishi, K., Katsunai, T., Taashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 403, 685-690 (2001)
 REFERENCE 5
 TITLE The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 1609)
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirao, K., Hirose, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

COMMENT

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komuro, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES

source

1. 1609
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:6430503015"
 /db_xref="WGI:2395676"
 /db_xref="taxon:10090"
 /clone="6430503015"
 /sex="male"
 /tissue_type="olfactory brain"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 330..1223
 /note="unnamed protein product; MSZF33 (FRAGMENT) homolog
 [Mus musculus] (SPR1088251, evidence: PASTY, 87.1%ID,
 100%length, match=255)
 putative"

CDS

/codon_start=1
 /protein_id="BAC27766.1"
 /db_xref="GI:26328051"
 /translation="MLETFRYNNANVGMEDSNIEHCSSRRHGRHRTGKPYE
 GTOGGAFTVHSHSLQWRKLIHTGKRYKNCQDKAVSRHSIIQIHRFTSGEPYCN
 QCGKAPVTHSHLKHVHTGTGKPYKNCQDKAFACNKLQKERTHTGKPYKQDC
 NNAFYESTLVQHKHTHTGKPYKNCQDKAFARHSHLKVHKIHTGKPYKNCQDC
 ALAYSTLVQHRTHTGKPYKNCQDCGKAFANQSYFVHKRIHTGKPYKDCGKAF
 VSSDILKRHRVHTGKPYKNCQDC"
 1594..1599
 /note="putative"
 1609
 /note="putative"
 /note="putative"

ORIGIN

Query Match 4.8%; Score 98.2; DB 11; Length 1609;
 Best Local Similarity 51.2%; Pred. No. 1.1e-05;
 Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

QY 779 AGCTCCCAAAAGTATATTGGACATATAGAGGTTGTAAGACCTATATGACCATC 838
 Db 521 AACTACTGAGAGAAACGCTACAAATGCAATGATGTATTAAGCTATTCACGACAG 580
 QY 839 ATTATTAGAGCAACATTTAAGAACCAACGATATGATGACCGTATTAATGTCAGTGA 898
 Db 581 CATTTACAAATACATTAAGAAACACATGTGGAGGAACCTATGAATGTA-----A 634
 QY 899 CGATTGTGTAAGAGATTTTTCAGAAATACATTTGGAAACATATGTAATGATCATTC 958
 Db 635 TCAATGTGTAAAGCCTTTACACACAGCATCTCAAAATACATATGTTATACATATAC 694

QY 1263 ACTGGTCAGTATTACATTTTCATATTAACAACCTGCATCCAAAACCTTAATGTCCTTAA 1322
 Db 1769 AAACCTTACTGAACATTAAGAAAATTTTCACTGAGAGAAACCTTA---TGAATGTGAAAAA 1825
 QY 1323 TGTGGTAAAGGTGTGTGGGAAAAAAGGTTTATCTTCACATATGTTTAAGCAGATGA 1381
 Db 1826 TGTGGCAAAAGCTTTTAACCACTCTCTCAATCTTACTAGACATTAAGAAAAGTCATACAGA 1884
 RESULT 5
 AUI23448 745 bp mRNA linear EST 01-AUG-2002
 LOCUS AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
 DEFINITION sequence.
 ACCESSION AUI23448
 VERSION AUI23448.1 GI:10948164
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 745)
 Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
 Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and
 Iwagai, T.
 HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,
 Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,
 Suzuki, Y., Sugano, S., Iwagai, T.)
 Unpublished (2000)
 CONTACT Contact: Takao Iwagai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 FEATURES
 source
 1..745
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RM2000319"
 /cell_type="leukocarcinoma"
 /cell_line="NT2"
 /clone_lib="NT2RM2"
 /note="Vector: PME18SFL3; mRNA from uninduced NT2 neuronal
 precursor cells"
 ORIGIN
 Query Match 4.7%; Score 97.6; DB 9; Length 745;
 Best Local Similarity 50.2%; Pired. No. 2.1e-05;
 Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

QY 966 AAACATTCATTTGTCAGTGTGTGTAAGGGGTATATTCGACACACTGGAAGA 1025
 Db 321 AAACCTCAAAATGTAAATAAATGTGAAAAGGCTTTAACGAGTGTGACACCTTACTCA 380
 QY 1026 CATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAAATGTCAAGACATTT 1085
 Db 381 CATAGGTAATTCATCTACTGAGAGAAACCTTACAAATGTGAAAATGTGAAAAGGCTTT 440
 QY 1086 TATTA---ACATCAATCTTTAAGCATCATATATATCTGTCATGAAAAACATTAAG 1142
 Db 441 AATCATTTCTGACACCTTACTACACATGAATATTCATCTGAGAGAAACCTTACAAA 500
 QY 1143 TGTAAACATATTAATAAGTTTTCATCTGACCTTCAAAATATAGACAAACATTAATAA 1202
 Db 501 TGTAAAGATGTGTAAAGCTTTTAAACACCTTTCAACCTTACTTAACATTAATTAATT 560
 QY 1203 CATCATGTGATCTCTGCTTCAATGATCATCATCTGATGTTGTTTAAATTTCCAA 1262
 Db 561 CATATCGAGAGAAAGCTTCAAAATGTAAAGAAATGTAAAGGCTTTTAAACCATCTCA 620
 QY 1263 ACTTGTCAGTATTACATTTTCATATTAACAACCTGCATCCAAAACCTTAATGTCCTTAA 1322
 Db 621 AAACCTTACTGAACATTAAGAAAATTTTCACTGAGAGAAACCTTA---TGAATGTGAAAAA 677
 QY 1323 TGTGGTAAAGGTGTGTGGGAAAAAAGGTTTATCTTCACATATGTTTAAGTCAT 1376
 Db 678 TGTGGCAAAAGCTTTTAACCACTCTCTCAATCTTACTAGACATTAAGAAAAGTCAT 731

RESULT 6
 LOCUS CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR08K10 of RPECI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL063921 GI:4941778
 VERSION AL063921.1
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazuo Oosawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPECI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR08K10"
 /clone_lib="RPECI-98"
 /note="end : TET3"

ORIGIN

QY 1028 TGAATCACCATTACAAAGTCATTAAATGACATTGGAAATTTGCAAGACATTTTA 1087
| | | | |
Db 1979 TAAGATTAATTCATCGAGAGAAACCATTAATGGAATGTGGCAAGCTTTTA 2038
| | | | |
QY 1088 TAAATCATCTTAAACATCATATATATCTGTCATGAAAAAATTAACGTG 1144
| | | | |
Db 2039 CCAATCTCAAACTTACCAACATTAAGAAAATTCATACGAGAAACCTTACATG 2098
| | | | |
QY 1145 TAAACATGTAAATAGTTTTCACCTGACCTTCAAAATTAAGACACATTAATTAACA 1204
| | | | |
Db 2099 TGAAGATGTGGCAAGCCTTTAACGATCTCTCAACCTTACTTAACATTAAGAAATTC 2158
| | | | |
QY 1205 TCATGTGATCTCT 1220
| | | | |
Db 2159 TACTGGAGAAAAAAGCT 2174
| | | | |

RESULT 8
A0005136 591 bp DNA linear GSS 27-JUN-1998
LOCUS
DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017,
genomic survey sequence.
ACCESSION A0005136
VERSION A0005136.1 GI:3082581
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,
AUTHORS Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other_GSSs: CIT-HSP-2290017.TR
JOURNAL
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..591
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7150844"
/db_xref="taxon:9606"
/clone="2290017"
/sex="Male"
/cell_type="Sperm"
/clone_1ib="CIT-HSP"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

ORIGIN
Query Match 4.5%; Score 92; DB 28; Length 591;
Best Local Similarity 52.2%; Pred. No. 0.00017;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

QY 726 GAAAGTACGAACCAATGATGATGATCTTTATCTCTCTTCATCAGCTCC 785
| | | | |
Db 69 GATGTGGCAAGCCTTTACCATCTTCAACCTTACTACACATTAAGTAATTCAT 128
| | | | |
QY 786 AAAAGTATTTGCAATATGAAGGTGTGATTAAGCCTATATATGACCATCATTA 845
| | | | |

Db 129 GGAGAGAGCCCTTCAATGTGAGAAATGTGTAAAGCTTTTAAACCACTTCAGCCCTT 188
| | | | |
QY 846 GAGCAACATTTTAAGAACCCAGATATGATTCGACCGTATTAATGTAACGTGCGATTGT 905
| | | | |
Db 189 ACTACATATTAAGCTTCTTCACTGTTTAAGAAAAAAGCTTACAAATGT-----GAAAGATGT 242
| | | | |
QY 906 GATTAAGCATTTTTCAGAAAATCAATTTGAAACACATATGATATCATCTTCGAAAAA 965
| | | | |
Db 243 GACAAAGCTTTTAACCGATTCATCTATCTTACTTAACATTAAGTAATTCATCTTCGAGAG 302
| | | | |
QY 966 AAACCATTCATTTGTCAGTGTGTGTAAGAGGCTTAATTCGACAAACCTTGAAGA 1025
| | | | |
Db 303 AAATCTTACAAATGTGTAACAAATGTGGCAAGGCTTTAACTGTCCTTCAACCTTACAAA 362
| | | | |
QY 1026 CATGAATACCCATCAAGATCATTTAATATGATCTTTGAATATGTCAGAAAGCATTT 1085
| | | | |
Db 363 CATTAAGAAATTTATCTGAGAGAAACCTTACAAATGTGAAGAAATGTGGCAAGCCTTT 422
| | | | |
QY 1086 TATTAACAT---CAATCTTAAAGACATCATATATATCTGTTGATGAAAAAATTAACG 1142
| | | | |
Db 423 AATGTCTTTCACACCTTACTACATTAAGATGATTCATATCTGAGAGAAACCTTACAA 482
| | | | |
QY 1143 TGTAAACATGTATATTAAGTTTCTACTGACCTTCAAAATTAAGCAACATTAATTA 1202
| | | | |
Db 483 TGTGAAGAAATGTGGCAAGCCTTTAACCACTCTCAAACTTACTATACATTAAGATTA 542
| | | | |
QY 1203 CATCATGTG 1212
| | | | |
Db 543 CATCTAGAG 552
| | | | |

RESULT 9
B99387 593 bp DNA linear GSS 26-JUN-1998
LOCUS
DEFINITION CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
genomic survey sequence.
ACCESSION B99387
VERSION B99387.1 GI:3027197
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,
AUTHORS Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..593
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7147044"
/db_xref="taxon:9606"
/clone="2281A10"
/sex="Male"
/cell_type="Sperm"
/clone_1ib="CIT-HSP"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.U., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantipod, S., Thomas, P.J., Touchman, J.W.,
 Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

FEATURES

source

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAK Plate: 65 Row: a Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 13386417
 This clone has the following problem: frame shifted.

Location/Qualifiers

```
1..2253
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C283H 11"
/db_xref="taxon:10090"
/clone="IMAGE:3674739"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(cloneal) outgrowth infected with the virus MMTV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DHIOB"
/notes="Vector: pCMV-SPORT6"
```

ORIGIN

Query Match 4.4%; Score 90.4; DB 11; Length 2253;
 Best Local Similarity 51.6%; Pred. No. 0.00013;
 Matches 261; Conservative 0; Mismatches 236; Indels 9; Gaps 2;

```

710 ACTTATTCATGAGTGAAGTGAAGCAACCAATGATATCATCTTTATATCTCTTC
810 ACCCTACAAATGTAATCATATGATTAAGCTTTTCAACACAGTTCCTTCAAACCTCA
770 TTCTTCATGAGTCCCAAAAAGTATTTGCAATATGAAGGGTGTGATTAAGCCTATTA
870 TAGAAGAACACATACCGGAAAGAAACCTTCAATGATATCATGATTAAGCCTTTTC
830 TCGACCATCATATTAAGACATTAAGAACCCACATATGATTCGACCGTATTAAG 889
930 CCGTAAACATTTGTGCAAAACCATATTAAGATCACTGAGAAAAACCTTCAAAAG 989
890 TACAGTGAAGATTTGATTAAGCAATTTTTCAGAAAATCAATTTGGAACACATATTT 949
990 TA-----ATCAATGATTAAGCCTTTTCAACAGACGACTCTTCAAACTCATTAAG 1043
950 ATCAATTCGGAAGAAAAACATTCATTTGTCAGTGTGTGTAAGGGGTTAATTTCTG 1009
1044 AACACATTCGGAAGAAACCTTCAATGATATCAATGATTAAGCCTTTTCTGTATA 1103
1010 ACAACATTTGAAGACATGAAGATCAACCATCAAAATCATTAATGATTAAGTTGAAA 1069
1104 ACATAGTCTCCAAACACATATGAAATACATCTGGGAAAAACCTTCAAAATGATATCA 1163
1070 TTGTCAAGAGCATTTTATTAACATCAATCTTTAAGACATCATATATATCTGTTCAT 1127
1164 ATGTATTAAGCCTTTTCAACACATTTCACTTCAAAATATATTAAGACATATCAAG 1223
```

QY 1128 -GAAAAACATTAACGTGTGAACATGATATTAAGTTTTCATCGACCTTCAAAATTAAGC 1186
 Db 1224 AGAGATGCTTCAAAATGATATGATGACAAAGCCTTTTACGATATGAAGATCTTCA 1283
 QY 1187 ACAACATTAATTAACATCAATGATG 1212
 Db 1284 GATTCAATGATTAAGAAACATATCTGAG 1309

RESULT 12
 BC022527
 LOCUS
 DEFINITION
 Homo sapiens hypothetical protein LOC148206, mRNA (cDNA clone
 IMAGE:4797857), containing frame-shift errors.
 BC022527
 BC022527.1 GI:18490307
 HTCC
 Homo sapiens (human)
 SOURCE
 ORGANISM

REFERENCE
AUTHORS

Strusberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Schmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buco, K.H., Scheffer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Datchenko, L., Marsina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Schetz, T.E., Brownstein, W.J., Usdin, T.B., Toshnyki, S.,
 Carninci, P., Prange, C., Rana, S.S., Lopoliano, N.A., Peters, G.J.,
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McMan, P.J.,
 McKernan, K.J., Malek, J.A., Gunatane, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Mizny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Witting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,
 Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 2694)
 Strusberg, R.
 Direct Submission
 Submitted (01-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshioyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAK Plate: 32 Row: f Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein
 Thus clone has the following problem: frame shifted.

OY	846	GAGCAACATTTTAAAGACCCACAGTAATGATGACCGTAAATGTAACAGTGGACGATTTGT	905
Db	1162	ACTACACATTAAGTTTCATTCATGTTTAAAGAAAAAACCCTTCAAAATG-----GAAAGAAATGT	1215
OY	906	GATTAAGCATTTTTCAGAAAAATTCACATTTGGAAAACACATATTGTATCATCTCCGAAAA	965
Db	1216	GACAAAGCTTTTAACCGAATCTCAATCTTACTTAAACCTTAAGATTAATTCATTCTTGGAGAG	1275
OY	966	AAACCATTCATTTGTCAGTGTGTGTGTAAGGGGTTAATTCTGACAAACACTTGAANA	1025
Db	1276	AAATCTTCAAAATGTGAACAATGTGGCAAAAGCTTTAACTGCTCTCAACCTTACAAAA	1335
OY	1026	CATGAATACACCCATCAACAAGTCATTTTAAATGTACTTTGAAAATGTGCAGAAAGCATTT	1085
Db	1336	CATGAACAATTTCAATCTGAGAGAAACCCCTACAAATGTGAAGAATGTGGCAAAAGCCTTT	1395
OY	1086	TATTAACAAT---CATCTTTAAGACATCATATTATCTGTTCATGAAAAAACAATTAAACG	1142
Db	1396	AATGTGCTTCACACCTTACTACACATTAAGATGATTCACTACTGGAGAGAAACCTTACAAA	1455
OY	1143	TGTAAACATGTAAATAAAGTTTCACTCGACCTTCAAAATTTAGCAACAACATTAATTAAAA	1202
Db	1456	TGTGAAGATGTGGCAAAAGCCTTTAACCACACTCTCAAAACTTACTATATCAATTAATATT	1515
OY	1203	CATCATGTGATCTCCT	1220
Db	1516	CATACTGGAGAAAACT	1533

/note="Origin: head neck; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cdna amplification were performed under low stringency conditions."

ORIGIN

Query Match	4.3%;	Score 89.2;	DB 10;	Length 604;
Best Local Similarity	53.7%;	Pred. No. 0.00044;		
Matches 234; Conservative	0;	Mismatches 193;	Indels 9;	Gaps 2

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butler, J.D., Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
 Scherch, A., Schein, J.E., Jones, S.J., and Marx, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 MEDLINE
 JOURNAL
 PUBMED
 REFERENCE
 2 (bases 1 to 3609)
 Scrausberg, R.
 Direct Submission
 Submitted (06-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
 Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgoun, C., Vogt, U.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNI at: <http://image.llnl.gov>
 Series: IRAK Plate: 69 Row: g Column: 20
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 13430873
 This clone has the following problem: frame shifted.
 Location/Qualifiers

1. 3609
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5502691"
 /tissue_type="Eye, retinoblastoma"
 /clone_id="NIH_MGC_67"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 4.3%; Score 89.2; DB 11; Length 3609;
 Best Local Similarity 53.7%; Pred. No. 0.00015;
 Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 788 AAAGTATATTGGCAATATGAGGGGTGATTAAGCCTATATATGACCATATTATTGA 847
 Db 587 AGAGAAACCTTACAAATGTGAGATGTGACAAAGTTTGTGCTAATACACACTTGA 646
 QY 848 GCACATTTTAAAGCCACAGTATGATCGACCGTATTAATGTACAGTGAAGATTTGA 907
 Db 647 AAGCATATGAGAAATTTCTACTGAGAGAAACCGTACAAATGTAAAGT-----TTGTGA 700
 QY 908 TAAAGCATTTTTCAGAAATCATTGGAACATATGATATCATATTCGAAAAA 967
 Db 701 CAAGCTTTTCAGAGTATTCACACCTGCGCAACATATTTATTCACACTGAGAGAA 760
 QY 968 ACCATTTCATTGTCAGTGTGTGTTAAAGGGGTTAATTTCTGACACACTGAAAAACA 1027
 Db 761 ACTTACAAAGTATATGATGTGTGCAAGACCTTTGTTCAAAATTCATCTCTTGAATGCA 820

QY 1028 TGAATTCACCATAC--AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAGCATT 1084
 Db 821 TAAGTCATTTCAATCTGAGAGAAATGTTCAAAGTAAATGAATGTGCAAGGTTTAA 880
 QY 1085 TTAATCAATCAATCTTTAAGACATCATATATTTATCTGTTCAATGAAAAAATTAACGTG 1144
 Db 881 TCAGAAATCAACCTTGCAATGTATCATATAGACTTCATCTGAGAGAAACCTTACAAAGTG 940
 QY 1145 TAAACATTTAATTAAGTTTTCACCTGACCTTCAAAATTTAGACACATTAATTTAAACA 1204
 Db 941 TAAATGAATGTGCGAAGGTTTAAATGAAAAATCAACCTTGACATCATATAGAGTTCA 1000
 QY 1205 TCATGTGATCTTCCT 1220
 Db 1001 TATCGAGAAAAAACCCT 1016

Search completed: July 28, 2004, 18:49:46
 Job time : 5418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:48:02 ; Search time 54 Seconds
(without alignments)
2155.733 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 412
Sequence: 1 MSESDETKSISLIISSSSSS.....PIWKARMDLPNETSVISR 412

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	412	100.0	412	3	AA93316
2	352	85.4	412	5	ABP73656
3	10	2.4	546	5	ABBS7056
4	9	2.2	58	4	ABG19930
5	9	2.2	114	4	AU30108
6	9	2.2	114	4	AU30387
7	9	2.2	128	3	AA857179
8	9	2.2	168	5	ABP64928
9	9	2.2	176	6	ADA55380
10	9	2.2	202	5	ABP28933
11	9	2.2	214	7	ADD47645
12	9	2.2	214	7	ADR60447
13	9	2.2	214	7	ADRS7980
14	9	2.2	214	7	ADRS7984
15	9	2.2	214	7	ADRS7984
16	9	2.2	214	7	ADRS7984
17	9	2.2	215	6	ABU07428
18	9	2.2	215	6	ABU07428
19	9	2.2	215	6	ABU07428
20	9	2.2	215	7	ADD40789
21	9	2.2	215	7	ADD40789
22	9	2.2	216	6	AA835866
23	9	2.2	220	6	ABG18955
24	9	2.2	223	4	AAU30109
25	9	2.2	226	4	ABH12356
			229	5	ABP64829

26	9	2.2	233	4	ABG19932	ABG19932	Novel hum
27	9	2.2	237	4	ABG09807	ABG09807	Novel hum
28	9	2.2	240	4	AU30388	AU30388	Novel hum
29	9	2.2	366	3	AA838042	AA838042	Arabidops
30	9	2.2	372	3	AA832376	AA832376	Arabidops
31	9	2.2	571	6	ABRS2911	ABRS2911	Protein s
32	9	2.2	589	7	ADBS8263	ADBS8263	Mouse RNA
33	9	2.2	699	7	AAV67598	AAV67598	Human adi
34	9	2.2	830	5	ABR04808	ABR04808	LDL recep
35	9	2.2	830	5	ABR04815	ABR04815	LDL recep
36	9	1.9	13	3	ABR01687	ABR01687	F1S1 poly
37	8	1.9	23	6	ABO12946	ABO12946	Mouse zin
38	8	1.9	23	6	ABO13305	ABO13305	Mouse zin
39	8	1.9	29	4	AAW2536	AAW2536	Peptide #
40	8	1.9	29	4	AAW2277	AAW2277	Human bon
41	8	1.9	29	4	ABG53963	ABG53963	Human liv
42	8	1.9	29	5	ABG42092	ABG42092	Human pep
43	8	1.9	63	4	AAW85534	AAW85534	Human imm
44	8	1.9	96	4	ABG27789	ABG27789	Novel hum
45	8	1.9	115	7	ADE58420	ADE58420	Rat Prote

ALIGNMENTS

```

RESULT 1
ID      AA93316 standard; protein, 412 AA.
XX
AC      AA93316;
XX
DT      04-SEP-2000 (first entry)
XX
DE      A transcription factor designated CATPITTA.
XX
KW      Transcription factor; CATPITTA; DNA-binding protein;
XX      ribosomal RNA 5S gene; fungal infection.
XX
OS      Candida albicans.
XX
FH      Key
FT      Misc-difference 193
FT      FT      /note= "Ser encoded by CTG"
FT      FT      /note= "Ser encoded by CTG"
FT      FT      /note= "Ser encoded by CTG"
XX
PN      WO200028037-A1.
XX
PD      18-MAY-2000.
XX
PF      09-NOV-1999; 99WO-FR002739.
XX
PR      10-NOV-1998; 98FR-00014147.
XX
PA      (HMRI ) HOECHST MARION ROUSSEL.
XX
PI      Bordon-Pallier F, Camier S, Sentenac A;
XX
PI      WPI, 2000-376549/32.
XX
PI      N-PSDB; AAA15398.
XX
DR      New nucleic acid encoding Candida albicans transcription factor, useful
XX      e.g. in screening for antineoplastic agents and for immunization.
XX
PS      Claim 12, Page 35-36; 45pp; French.
XX
XX
XX      The present sequence represents a Candida albicans transcription factor,
XX      designated CATPITTA. The polypeptide is a DNA-binding protein, which is
XX      involved in initiating transcription of the ribosomal RNA 5S gene. The
XX      polynucleotide is used to screen for its specific inhibitors, potentially
XX      useful as antineoplastic agents, to raise an antibody response that is
XX      protective against fungal infection and to raise antibodies. Such
XX      antibodies, as well as the polypeptides and polynucleotides are used in
  
```

CC compositions for diagnosing and treating fungal infections, e.g. by
 CC detecting polymorphisms and mutations

XX Sequence 412 AA;

Query Match 100.0%; Score 412; DB 3; Length 412;
 Best Local Similarity 100.0%; Pred. No.0;
 Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSESDETKSISLSSSSSRPKYICTYEGCDKAVNRPSELBOHLRTHSNDRPYKCTVD 60
 DB 1 MSESDETKSISLSSSSSRPKYICTYEGCDKAVNRPSELBOHLRTHSNDRPYKCTVD 60
 QY 61 DCDKAFPRKSHLETHIVSHSEKRPFHCSVCGKGVNSRQHLKRHEITHTKSFCTFENCQE 120
 DB 61 DCDKAFPRKSHLETHIVSHSEKRPFHCSVCGKGVNSRQHLKRHEITHTKSFCTFENCQE 120
 QY 121 AFYKHSIRHILISVHEKTLTKQCNKVFTRPSTLAQHLKHGGSFAYQCDHGGCTKNF 180
 DB 121 AFYKHSIRHILISVHEKTLTKQCNKVFTRPSTLAQHLKHGGSFAYQCDHGGCTKNF 180
 QY 181 QTWSVLQFHIKQSHPKLCKPCGCGVGGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 DB 181 QTWSVLQFHIKQSHPKLCKPCGCGVGGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 QY 241 KNEIVHEYNIFHDGNITPDDLKETEYVKLENLDQGSKLNHLHLETEKLVBEDEBDEE 300
 DB 241 KNEIVHEYNIFHDGNITPDDLKETEYVKLENLDQGSKLNHLHLETEKLVBEDEBDEE 300
 QY 301 DSLDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPKXNCDMFSEKREY 360
 DB 301 DSLDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPKXNCDMFSEKREY 360
 QY 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
 DB 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412

RESULT 2
 ABP73656
 ID 7 ABP73656 standard; protein; 412 AA.

XX AC ABP73656;

XX DT 30-JAN-2003 (first entry)

XX DE Candida albicans essential protein SEQ ID NO 7493.

XX KM Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KM signal transduction; DNA replication; cell division; growth;
 KM proliferation; Candida albicans; fungicide; antifungal.

XX OS Candida albicans.

XX PN WO200253728-A2.

XX PD 11-JUL-2002.

XX PF 26-DEC-2001; 2001WO-US049486.

XX PR 29-DEC-2000; 2000US-0259128P.

XX PR 20-FEB-2001; 2001US-00792024.

XX PR 22-AUG-2001; 2001US-0314050P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;

XX DR WI; 2002-566694/60.

XX DR N-PSDB; ABZ32206.

XX PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.
 XX Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.

XX PS The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans protein used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office

XX SQ Sequence 412 AA;

Query Match 85.4%; Score 352; DB 5; Length 412;
 Best Local Similarity 100.0%; Pred. No.0;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 DCDKAFPRKSHLETHIVSHSEKRPFHCSVCGKGVNSRQHLKRHEITHTKSFCTFENCQE 120
 DB 61 DCDKAFPRKSHLETHIVSHSEKRPFHCSVCGKGVNSRQHLKRHEITHTKSFCTFENCQE 120
 QY 121 AFYKHSIRHILISVHEKTLTKQCNKVFTRPSTLAQHLKHGGSFAYQCDHGGCTKNF 180
 DB 121 AFYKHSIRHILISVHEKTLTKQCNKVFTRPSTLAQHLKHGGSFAYQCDHGGCTKNF 180
 QY 181 QTWSVLQFHIKQSHPKLCKPCGCGVGGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 DB 181 QTWSVLQFHIKQSHPKLCKPCGCGVGGKGLSSHMLSHDSTMIKIMTCDYCDVGKFAK 240
 QY 241 KNEIVHEYNIFHDGNITPDDLKETEYVKLENLDQGSKLNHLHLETEKLVBEDEBDEE 300
 DB 241 KNEIVHEYNIFHDGNITPDDLKETEYVKLENLDQGSKLNHLHLETEKLVBEDEBDEE 300
 QY 301 DSLDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPKXNCDMFSEKREY 360
 DB 301 DSLDEKSDVRSMSAQRSTKSTFTASLEGSYSKILNSGKKINCPKXNCDMFSEKREY 360
 QY 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
 DB 361 DLRRHLKMHNDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412

RESULT 3
 ABB57056
 ID ABB57056 standard; protein; 546 AA.

XX AC ABB57056;

XX DT 07-MAR-2002 (first entry)

XX DE Mouse ischaemic condition related protein sequence SEQ ID NO:106.

XX KM Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KM vasospastic ischaemia; ischaemic condition; ischaemic disease.

XX OS Mus musculus.
XX PA WO200188188-A2.
XX PN 22-NOV-2001.
XX PD 18-MAY-2001; 2001WO-JP004192.
XX PF 18-MAY-2000; 2000JP-00145977.
XX PR 18-MAY-2000; 2000JP-00145977.
XX PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX DR WPI; 2002-034733/04.
XX DR N-PSDB; AB199270.
XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or by
PT determining the expression profile of a gene group comprising these
PT genes.
XX PS Claim 2; Page 311-313; 2690pp; English.
XX CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (i) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (i). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischemia, occlusive ischaemia or vasospastic ischaemia) by measuring the
CC expression levels of particular genes (AB199202 to AB199912, encoding the
CC protein sequences in AB199202 to AB199912) or by determining the
CC expression profile of a gene group comprising these genes. The expression
CC levels or expression profiles produced by these genes are used as an
CC indicator when screening for ischaemic condition-improving drugs or
CC therapeutics for ischaemic diseases. AB199913 and AB199914 represent PCR
CC primers for a mouse ischaemic condition related sequence, which are used
CC in the exemplification of the present invention
XX SQ Sequence 546 AA;
XX
XX Query Match 2.4%; Score 10; DB 5; Length 546;
XX Best Local Similarity 100.0%; Pred. No. 1.4;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 100 LKRRHITHTK 109
DB 505 LKRRHITHTK 514
RESULT 4
ABG19930
ID ABG19930 standard; protein; 58 AA.
XX AC ABG19930;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #19921.
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM Food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS84117.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX PS Claim 20; SEQ ID NO 50289; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AB000010-AB030377 represent novel human diagnostic
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 58 AA;
XX
XX Query Match 2.2%; Score 9; DB 4; Length 58;
XX Best Local Similarity 100.0%; Pred. No. 1.7;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EEDDEDEED 301
DB 31 EEDDEDEED 39
RESULT 5
AAU30108
ID AAU30108 standard; protein; 114 AA.
XX AC AAU30108;
XX DT 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #599.
XX KM Human; vaccination; gene therapy; nutritional supplement;
XX KM stem cell proliferation; hematopoiesis; nerve tissue regeneration;
XX KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US008656.
XX PR 18-APR-2000; 2000US-00552929.
XX PR 26-JAN-2001; 2001US-00770160.

XX (HXSE-) HXSEQ INC.
 XX
 XX Tang YT, Liu C, Dimanac RT,
 XX
 XX WPI, 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 XX
 XX
 XX Claim 20; Page 241; 765bp; English.
 XX
 XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAD29510-AAD33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 XX Sequence 114 AA;
 XX

Query Match	2.2%	Score 9;	DB 4;	Length 114;
Best Local Similarity	100.0%	Pred. No. 3.1;		
Matches	9;	Conservative	0;	Mismatches 0; Gaps 0;
Oy	293	EDDEDDED	301	
Db	86	BEDDEDDED	94	
RESULT 6				
AAU30387				
ID	AAU30387	standard; protein; 114 AA.		
XX	AAU30387;			
AC				
DT	18-DEC-2001	(first entry)		
XX				
DE	Novel human secreted protein #878.			
XX				
KM	Human; vaccination; gene therapy; nutritional supplement;			
KW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;			
XX	immune suppression; immune stimulation; anti-inflammatory; leukaemia.			
OS	Homo sapiens.			
XX				
PN	WO200179449-A2.			
XX				
PD	25-OCT-2001.			
XX				
PF	16-APR-2001; 2001WO-US008656.			
XX				
PR	18-APR-2000; 2000US-00552929.			
PR	26-JAN-2001; 2001US-00770160.			
XX				
XX				
PA	(HYSB-) HYSBQ INC.			
XX				
PI	Tang YT, Liu C, Drmanac RT;			
XX				
XX	WPI; 2001-611725/70.			
PT	Nucleic acids encoding a range of human polypeptides, useful in genetic			

FT	vaccination, testing and therapy.
XX	
PS	Claim 20, Page 287, 765pp; English.
XX	
CC	The invention relates to novel human secreted polypeptides. The
CC	polypeptides and antibodies to the polypeptides are useful for
CC	determining the presence of or predisposition to a disease associated
CC	with altered levels of polypeptide. The polypeptides are also useful for
CC	identifying agents (agonists and antagonists) that bind to them. Cells
CC	expressing the proteins are useful for identifying a therapeutic agent
CC	for use in treatment of a pathology related to aberrant expression or
CC	physiological interactions of the polypeptide. Vectors comprising the
CC	nucleic acids encoding the polypeptides and cells genetically engineered
CC	to express them are also useful for producing the proteins. The proteins
CC	are useful in genetic vaccination, testing and therapy, and can be used
CC	as nutritional supplements. They may be used to increase stem cell
CC	proliferation, to regulate haematopoiesis; and in bone, cartilage, tendon
CC	and/or nerve tissue growth or regeneration; immune suppression and/or
CC	stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC	AU929510-AU933304 represent the amino acid sequences of novel human
CC	secreted proteins of the invention
XX	
SQ	Sequence 114 AA;
	Query Match 2.2%; Score 9; DB 4; Length 114;
	Best Local Similarity 100.0%; Pred. No. 3.1;
	Matches 9; Conservative 0; Mismatches 0; Gaps 0;
QY	293 EDEDEDPD 301
DB	86 EDEDEDPD 94

RESULT 7	
AAB57179	
ID	AAB57179 standard; protein, 128 AA.
XX	
AC	AAB57179;
XX	
DT	13-MAR-2001 (first entry)
XX	
DE	Human prostate cancer antigen protein sequence SEQ ID NO:1757.
XX	
KW	Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW	neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular;
KW	vulnereary; gastrointestinal; nephrotoptic; antinfecitive; gynaecological;
KW	antibacterial; gene therapy; neural; immune; reproductive; renal;
KW	gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW	wound; infectious disease.
XX	
OS	Homo sapiens.
XX	
PN	WO20005174-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US005988.
XX	
PR	12-MAR-1999; 99US-0124270P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PA	(ROSE/) ROSEN C A.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-587513/55.
DR	N-PSDB; AAF16382.
XX	
PT	Prostate cancer associated gene sequences, referred to as prostate cancer
PT	antigens, useful for treatment, prevention, and diagnosis of disorders
PT	such as prostate cancer.
XX	
PS	Claim 11; Page 2225-2226; 2338pp; English.

XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnerrary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention
CC
SQ Sequence 128 AA;
QY
Best Match 2.2%; Score 9; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 293 EEDDEDEED 301
105 EEDDEDEED 113

RESULT 8
ABP64928
ID ABP64928 standard; protein; 168 AA.
XX
AC ABP64928;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human protein SEQ ID 588.
XX
KW Human; expressed sequence tag; EST; haematopoietic disorder;
KW central nervous system disease; viral infection;
KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; neurotropic;
KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective.
XX
OS Homo sapiens.
XX
PN WO200259260-A2.
XX
PD 01-AUG-2002.
XX
PF 16-NOV-2001; 2001WO-US042950.
XX
PR 17-NOV-2000; 2000US-00714936.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-590824/63.
DR N-PSDB; ABQ99514.
XX
PS Claim 20; SEQ ID NO 588; 394bp; English.
CC The present invention relates to novel human coding sequences (ABQ99268-
CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC therapeutic, diagnostic and research methods. The polynucleotides may be

CC used in the field of molecular biology as hybridisation probes, primers
CC for PCR, for chromosome and gene mapping, for the recombinant production
CC of protein, or in generation of anti-gene DNA or RNA. The
CC polynucleotides are useful in diagnostics as expressed sequence tags
CC (ESTs) for identifying expressed genes or for physical mapping of the
CC human genome. The proteins may be used as molecular weight markers, or as
CC nutritional sources or supplements. The proteins may be used to maintain
CC and expand cell population in a totipotent or pluripotent state
CC useful for re-engineering damaged or diseased tissues, transplantation,
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC polynucleotides and proteins are useful for preventing, treating or
CC ameliorating disorders involving aberrant protein expression or
CC biological activity, e.g. haematopoietic disorders, central/peripheral
CC nervous system diseases, mechanical and traumatic disorders, non-healing
CC wounds, immune deficiencies and disorders, infectious diseases caused by
CC viral, bacterial or fungal infection, autoimmune disorders, allergic
CC reactions and conditions, coagulation disorders, or cancer. The
CC polynucleotide sequences of the invention were assembled from ESTs
CC isolated mainly by sequencing by hybridisation, and in some cases,
CC sequences obtained from one or more public databases. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIP0 at
CC ftp.wipo.int/pub/published_pct_sequences
CC
SQ Sequence 168 AA;
QY
Best Match 2.2%; Score 9; DB 5; Length 168;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 293 EEDDEDEED 301
142 EEDDEDEED 150

RESULT 9
ADA55380
ID ADA55380 standard; protein; 176 AA.
XX
AC ADA55380;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 2948.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Neurotropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
XX
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakaratsu A, Sato H, Ishii S;
PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR N-PSDB; ADA53741.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2948; 2055p; English.
PS
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 176 AA;
SQ
Query Match 2.2%; Score 9; DB 6; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EDEDEDEED 301
DB 149 EDEDEDEED 157
RESULT 10
ID ABP28933 standard; protein; 202 AA.
AC ABP28933;
XX
XX 02-JUL-2002 (first entry)
DT
XX Streptococcus polypeptide SEQ ID NO 7042.
DE
XX Streptococcus pyogenes.
OS
XX Streptococcus pyogenes.
XX
XX WO200234771-A2.
PN
XX
XX 02-MAY-2002.
PD
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX
XX 24-NOV-2000; 2000GB-00028727.
XX
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettein H;
XX
XX WPI; 2002-352536/38.
XX
XX N-PSDB; ABN69564.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3865; 4525p; English.
PS
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX Sequence 202 AA;
SQ
Query Match 2.2%; Score 9; DB 5; Length 202;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EDEDEDEED 301
DB 189 EDEDEDEED 197
RESULT 11
ID ADD47645 standard; protein; 214 AA.
AC ADD47645;
XX
XX 29-JAN-2004 (first entry)
DT
XX Human Protein P09429, SEQ ID NO 13341.
DE
XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SN1; Chung.
XX
XX Homo sapiens.
XX
XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
XX
XX 01-NOV-2001; 2001US-0346382P.
XX
XX 26-NOV-2001; 2001US-0333347P.
XX
XX (GENO) GEN HOSPITAL CORP.
XX
XX (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
PI
XX
XX WPI; 2003-268312/26.
XX
XX GENBANK; P09429.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017p; English.
PS
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 214 AA:

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
DB 187 EDEDEDEED 195

RESULT 12
ADE60447

ID ADE60447 standard; protein; 214 AA.

XX ADE60447;

DT 29-JAN-2004 (first entry)

DE Human Protein P09429, SEQ ID NO 6356.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

DR GENBANK; P09429.

XX
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.

XX
XX PS Claim 1; Page; 1017pp; English.

XX
XX CC The invention discloses a composition comprising two or more isolated rat
XX CC or human polynucleotides or a polynucleotide which represents a fragment,
XX CC derivative or allelic variation of the nucleic acid sequence. Also
XX CC claimed are a vector comprising the novel polynucleotide, a host cell
XX CC comprising the vector, a method for identifying a nucleotide sequence
XX CC which is differentially regulated in an animal subjected to pain and a
XX CC kit to perform the method, an array, a method for identifying an agent
XX CC that increases or decreases the expression of the polynucleotide sequence
XX CC that is differentially expressed in neuronal tissue of a first animal
XX CC subjected to pain, a method for identifying a compound which regulates
XX CC the expression of a polynucleotide sequence which is differentially
XX CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 214 AA:

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
DB 187 EDEDEDEED 195

RESULT 13
ADE57980

ID ADE57980 standard; protein; 214 AA.

XX ADE57980;

DT 29-JAN-2004 (first entry)

DE Human Protein P09429, SEQ ID NO 3848.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

DR GENBANK; P09429.

XX
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.

XX
XX PS Claim 1; Page; 1017pp; English.

XX
XX CC The invention discloses a composition comprising two or more isolated rat
XX CC or human polynucleotides or a polynucleotide which represents a fragment,
XX CC derivative or allelic variation of the nucleic acid sequence. Also
XX CC claimed are a vector comprising the novel polynucleotide, a host cell
XX CC comprising the vector, a method for identifying a nucleotide sequence
XX CC which is differentially regulated in an animal subjected to pain and a
XX CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 214 AA;

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 187 EDEDEDEED 195

RESULT 14

ADBE57984
ID ADE57984 standard; protein; 214 AA.

XX ADE57984;

DT 29-JAN-2004 (first entry)

DE Human Protein P09429, SEQ ID NO 3852.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.

XX Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

XX GENBANK; P09429.

XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017p; English.
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 214 AA;

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 187 EDEDEDEED 195

RESULT 15

ADBE60732
ID ADE60732 standard; protein; 214 AA.

XX ADE60732;

DT 29-JAN-2004 (first entry)

DE Human Protein P09429, SEQ ID NO 6644.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.

XX Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

XX GENBANK; P09429.

XX New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page: 1017pp; English.
PS
XX

CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 214 AA;

Query Match 2.2%; Score 9; DB 7; Length 214;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EEDDEDEED 301
Db 187 EEDDEDEED 195

Search completed: July 22, 2004, 16:54:15
Job time : 56 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:48:32 ; Search time 13 Seconds

(Without alignments)
1650.223 Million cell updates/sec

Title: US-09-831-804-3

Sequence: 1 MSSEDTKTSISSISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	14	3.4	TF3A_SCHPO	Q9ut15 schizosacch
2	10	2.4	2235_MOUSE	Q61116 mus musculu
3	9	2.2	RPOB_STRPB	Q8nrf7 streptococc
4	9	2.2	RPOB_STRPY	P58053 streptococc
5	9	2.2	HMIX_HUMAN	Q9ugv6 homo sapien
6	9	2.2	HMGI_BOVIN	P10103 bos taurus
7	9	2.2	HMGI_HUMAN	P09429 homo sapien
8	9	2.2	214	P12682 sus scrofa
9	9	2.2	KC2B_NEUCR	Q8c912 neuropepti
10	9	2.2	TRT_DROME	P19351 drosophila
11	9	2.2	YK18_YEAST	P36076 saccharomyc
12	9	2.2	RGPI_MOUSE	P46061 mus musculu
13	9	2.2	CENB_CRIGR	P48988 cricetulus
14	9	2.2	ECM2_HUMAN	P34440 plasmodium
15	9	2.2	CH60_PIARG	Q99299 saccharomyc
16	9	2.2	YPS8_YEAST	Q61362 rattus norv
17	9	2.2	YIP2_MOUSE	Q9e9e9 mus musculu
18	8	1.9	HSP3_RAT	Q63656 rattus norv
19	8	1.9	NUFM_RAT	P40620 vicia faba
20	8	1.9	HMGI_YTCFA	P07156 cricetulus
21	8	1.9	HMGI_CRIGR	P07155 mus musculu
22	8	1.9	HMGI_MOUSE	P87179 schizosacch
23	8	1.9	YB1E_SCHPO	Q64953 eucalyptus
24	8	1.9	FLI_EDOGL	P24827 pseudorabie
25	8	1.9	IE68_PVKKA	Q75381 homo sapien
26	8	1.9	PEXE_HUMAN	Q35738 mus musculu
27	8	1.9	KLFC_MOUSE	Q26486 spodoptera
28	8	1.9	FXM4_SPOFR	P27797 homo sapien
29	8	1.9	CRTC_HUMAN	Q46107 drosophila
30	8	1.9	LIP1_DROME	Q9p677 homo sapien
31	8	1.9	IRX2_HUMAN	P81066 mus musculu
32	8	1.9	IRX2_MOUSE	Q00294 homo sapien
33	8	1.9	TDL1_HUMAN	

34	8	1.9	547	1	IF37_MOUSE	Q70194 mus musculu
35	8	1.9	584	1	YMB3_YEAST	Q04228 saccharomyc
36	8	1.9	591	1	PAK4_HUMAN	Q96013 homo sapien
37	8	1.9	628	1	DY11_MOUSE	Q88485 mus musculu
38	8	1.9	628	1	HAP1_MOUSE	Q35668 mus musculu
39	8	1.9	629	1	HAP1_RAT	P54256 rattus norv
40	8	1.9	643	1	DY11_RAT	Q63100 rattus norv
41	8	1.9	645	1	DY11_HUMAN	Q14576 homo sapien
42	8	1.9	689	1	MEA_ARATH	Q65312 arabidopsis
43	8	1.9	694	1	NUCLE_CHICK	P15771 galus galli
44	8	1.9	695	1	APP2_MOUSE	Q06335 mus musculu
45	8	1.9	701	1	UBF2_XENILA	P25980 xenopus lae

ALIGNMENTS

RESULT 1
TF3A_SCHPO STANDARD; PRT; 374 AA.
ID O9UT15;
AC 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
GN Transcription factor IIA (Factor A) (TFIIIA).
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
RX SEQUENCE FROM N.A., AND FUNCTION.
RP MEDLINE=22082319; PubMed=12087160;
RA Schulman D.B., Seczer D.R.;
RT Identification and characterization of transcription factor IIA from
RL Schizosaccharomyces pombe.";
RL Nucleic Acids Res. 30:2772-2781(2002).
RN
RP SOURCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21648401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Vancleert G., Aert R., Robben J., Grymopier B.,
RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Muehler-Auer S.,
RA Gabell C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirauch H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffinet A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Roquet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado I., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Renvelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
-!- FUNCTION: Is required for correct transcription of 5S RNA genes by
RNA polymerase III. Also binds the transcribed 5S RNA's. Initiates
transcription of the 5S ribosomal RNA gene.
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AY091590; AAM00046.1; -
 DR EMBL: AL132675; CAB59689.1; -
 DR PIR: T37676; T37676.
 DR HSSP: P07248; IARD.
 DR GeneDB: SPombe; SPAC14.09c; -
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 10.
 DR ProDom: PD000003; Znf_C2H2; 1.
 DR SMART: SM00355; Znf_C2H2; 9.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 9.
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 23 47 C2H2-TYPE.
 FT ZN_FING 53 77 C2H2-TYPE.
 FT ZN_FING 83 107 C2H2-TYPE.
 FT ZN_FING 113 138 C2H2-TYPE.
 FT ZN_FING 144 169 C2H2-TYPE.
 FT ZN_FING 204 226 C2H2-TYPE.
 FT ZN_FING 236 261 C2H2-TYPE.
 FT ZN_FING 267 291 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 349 374 C2H2-TYPE.
 SQ SEQUENCE 374 AA; 43851 MM; 7469C701FF08FF6 CRC64;

Query Match 3.4%; Score 14; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 RPSLLEQHLRTHSN 51
 DB 36 RPSLLEQHLRTHSN 49

RESULT 2
 Z235_MOUSE STANDARD; PRT; 645 AA.
 AC Q61116; Q8CE70;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 235 (Zinc finger protein 93) (Zfp-93).
 GN ZNF235 OR ZFP93.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Testis;
 RX MEDLINE=96207307; PubMed=8617494;
 RA Shannon M., Ashworth L.K., Mucenski M.L., Lamerdin J.E., Branscomb E.,
 RA Stubbs L.;
 RT "Comparative analysis of a conserved zinc finger gene cluster on human
 RT chromosome 19q and mouse chromosome 7.";
 RL Genomics 33:112-120(1996).
 RU [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Skin, and Thymus;
 RX MEDLINE=22354683; PubMed=1246851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito K., Suzuki H., Yamakata I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schmitt L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,

RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi J., Godzik A., Gough J.,
 RA Grimmond S., Gusttinch S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedziński R.M., King B.L.,
 RA Kanojaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pereira G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszaw-Boris A., Yanagisawa M., Yang L., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirokane Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh N., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney R., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RT [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusta K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huljk S.W.,
 RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A.C., Shcherbenko Y., Bonfield G.G.,
 RA Whiting M., Madan A., Young A.C., Shcherbenko Y., Bonfield G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May function as a transcription factor.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U46186; AAB03529.1; -
 DR EMBL: AK028905; BAC26187.1; -
 DR EMBL: AK040325; BAC30565.1; -
 DR EMBL: BC003776; AAB03776.1; -
 DR HSSP: P25490; IZNM.
 DR MGD: MGI:107611; Zfp93.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR007087; Znf_C2H2.
 DR InterPro: IPR007086; Znf_C2H2_sub.
 DR Pfam: PF01352; KRAB.1.
 DR Pfam: PF00096; zf-C2H2; 13.
 DR PRINTS: PR00048; ZINCFINGER.

```
DR ProDom: PD000003; Znf_C2H2; 11.
DR SMART; SM00349; KRAb_1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS50805; KRAb; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 8
FT ZN_FING 285 307 KRAb.
FT ZN_FING 313 335 C2H2-TYPE.
FT ZN_FING 341 363 C2H2-TYPE.
FT ZN_FING 369 391 C2H2-TYPE.
FT ZN_FING 397 419 C2H2-TYPE.
FT ZN_FING 425 447 C2H2-TYPE.
FT ZN_FING 453 475 C2H2-TYPE.
FT ZN_FING 481 503 C2H2-TYPE.
FT ZN_FING 509 531 C2H2-TYPE.
FT ZN_FING 537 559 C2H2-TYPE.
FT ZN_FING 565 587 C2H2-TYPE.
FT ZN_FING 593 615 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.
FT CONFLICT 85 85 A -> P (IN REF. 2; BAC26187).
SQ SEQUENCE 645 AA; 73012 MM; 7459CD140F5AF469 CRC64;

Query Match 2.4%; Score 10; DB 1; Length 645;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KPFGCVCGK 92
Db 451 KPFGCVCGK 460

RESULT 3
RPOE_STRP8
ID RPOE_STRP8 STANDARD; PRT; 191 AA.
AC GBNZF7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable DNA-directed RNA polymerase delta subunit (RNAP delta
factor).
GN RPOB OR SPY18_1960.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy D.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
group A Streptococcus strains associated with acute rheumatic fever
outbreaks.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
RL [3]
CC -!- FUNCTION: Participates in both the initiation and recycling phases
of transcription. In the presence of the delta subunit, RNAP
displays an increased specificity of transcription, a decreased
affinity for nucleic acids, and an increased efficiency of RNA
synthesis because of enhanced recycling (by similarity).
CC -!- SUBUNIT: RNAP is composed of a core of 2 alpha, a beta and a beta'
subunits. The core is associated with a delta subunit and one of
several sigma factors (by similarity).
CC -!- SIMILARITY: Belongs to the rpoB family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010100; AAL98451.1; -.
DR HAMAP; MF 00357; -; 1.
DR InterPro; IPR007759; RNA pol delta.
DR Pfam; PF05066; RNA pol delta_1.
KW DNA-directed RNA polymerase; Transcription; Complete proteome.
FT DOMAIN 106 191 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 191 AA; 22250 MM; CD081EFA0BC7C329 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EEDDEDEED 301
Db 178 EEDDEDEED 186

RESULT 4
RPOE_STRPY
ID RPOE_STRPY STANDARD; PRT; 191 AA.
AC P58053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable DNA-directed RNA polymerase delta subunit (RNAP delta
factor).
GN RPOE OR SPY1895 OR SPY13 1633 OR SPS0233.
OS Streptococcus pyogenes, and
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phase-encoded toxins, the high-virulence phenotype, and clone
emergence.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RL [3]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RT Genome Res. 13:1042-1055(2003).
RL [3]
CC -!- FUNCTION: Participates in both the initiation and recycling phases
of transcription. In the presence of the delta subunit, RNAP
displays an increased specificity of transcription, a decreased
```

CC affinity for nucleic acids, and an increased efficiency of RNA
 CC synthesis because of enhanced recycling (by similarity).
 CC -1- SUBUNIT: RNP is composed of a core of 2 alpha, a beta and a beta'
 CC subunits. The core is associated with a delta subunit and one of
 CC several sigma factors (by similarity).
 CC -1- SIMILARITY: Belongs to the rpoC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB006614; AAK34603.1; -
 CC EMBL; AB014167; AAM80240.1; ALT_INIT.
 CC EMBL; AP005141; BAC63328.1; ALT_INIT.
 CC HAMAP; MF_00357; -; 1.
 CC InterPro; IPR007759; RNA_pol_delta.
 CC Pfam; PF05066; RNA_pol_delta; 1.
 CC DNA-directed RNA polymerase; Transcription; Complete proteome.
 CC FT DOMAIN 118 191 ASP/GLU-RICH (ACIDIC).
 CC SEQUENCE 191 AA; 22264 MW; D1781EFA0BC6926C CRC64;
 CC -----
 CC Query March 2.2%; Score 9; DB 1; Length 191;
 CC Best Local Similarity 100.0%; Pred. No. 0.49; Mismatches 0; Gaps 0;
 CC Matches 9; Conservative 0; Indels 0;
 CC -----
 CC 293 EDEDEDED 301
 CC |||||
 CC Db 178 EDEDEDED 186
 CC -----
 CC RESULT 5
 CC HMLX HUMAN STANDARD; PRT; 211 AA.
 CC ID HMLX_HUMAN
 CC AC O9UGV6;
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE High mobility group protein 1-like 10 (HMG-1L10).
 CC GN HMG1L10.
 CC OS Homo sapiens (Human).
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC OX NCBI_TaxID=9606;
 CC RX SEQUENCE FROM N.A.
 CC MEDLINE=20057165; PubMed=10591208;
 CC Dunham I., Hunt A.R., Collins J.E., Brunikewich R., Beare D.M.,
 CC Clump M., Smink L.U., Ainscough R., Almeida J.P., Babbage A.K.,
 CC Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 CC Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 CC Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 CC Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 CC Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson R.,
 CC Dharmi P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 CC Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 CC Gilbert J.G.R., Goward M.E., Grahame D.V., Griffiths M.N.D., Hall C.,
 CC Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 CC Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 CC Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 CC Matryn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 CC McElay J., McLaren S., McMurtry A.A., Milne S.A., Mortimore B.J.C.T.,
 CC Ogden C.N., Pavlitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 CC Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Rose M.T.,
 CC Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 CC Soderlund C., Spraggon I., Steward C.A., Suleston J.E., Swann R.M.,
 CC Vaddin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 CC Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 CC Wright C.L., Hubbard K., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 CC Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,

RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuayama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kerton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissos S., Murray J., Miller N., Mux P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
 RA Schuet P., Walker C., Wamsley A., Wohlmann P., Fepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saito S.,
 RA Badar M.L., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shibuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Pransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tlilhan Y., Wright H.,
 RT "The DNA sequence of human chromosome 22."
 RL Nature 402:489-495(1999).
 CC -1- FUNCTION: Binds preferentially single-stranded DNA and unwinds
 CC double stranded DNA (by similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
 CC -1- SIMILARITY: Contains 2 HMG box domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z95115; CAB62951.1; -
 CC HSSP; P07156; HNHN.
 CC Genew; HGNC:494; HMG1L10.
 CC InterPro; IPR00135; Highmobility_12.
 CC InterPro; IPR00910; HMG_12_box.
 CC Pfam; PF00505; HMG_box; 2.
 CC PRINTS; PR00866; HIGHMOBILITY12.
 CC SMART; SM00398; HMG; 2.
 CC PROSITE; PS00353; HMG_BOX_1; 1.
 CC PROSITE; PS50118; HMG_BOX_2; 2.
 CC KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
 CC FT DNA_BIND 9 79 HMG_BOX 1.
 CC FT DNA_BIND 95 163 HMG_BOX 2.
 CC FT DOMAIN 186 211 ASP/GLU-RICH (ACIDIC).
 CC SEQUENCE 211 AA; 24218 MW; 2A53BA2A6A6DF7CD CRC64;
 CC -----
 CC Query March 2.2%; Score 9; DB 1; Length 211;
 CC Best Local Similarity 100.0%; Pred. No. 0.53; Mismatches 0; Gaps 0;
 CC Matches 9; Conservative 0; Indels 0;
 CC -----
 CC 293 EDEDEDED 301
 CC |||||
 CC Db 188 EDEDEDED 196
 CC -----
 CC RESULT 6
 CC HMG1 BOVIN STANDARD; PRT; 214 AA.
 CC ID HMG1_BOVIN
 CC AC P10103;
 CC DT 01-MAR-1989 (Rel. 10, Last sequence update)
 CC DT 01-AUG-1990 (Rel. 15, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE High mobility group protein 1 (HMG-1).
 CC GN HMG1 OR HMG1.
 CC OS Bos taurus (Bovine).
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 CC Bovidae; Bovinae; Bos.
 CC OX NCBI_TaxID=9913;
 CC RX SEQUENCE FROM N.A.

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holstein; TISSUE=Fetal thymus;
RX MEDLINE=89057489; PubMed=3194213;
RA Kaplan D.J., Duncan C.H.;
RT "Full length cDNA sequence for bovine high mobility group 1 (HMG1) protein."
RL Nucleic Acids Res. 16:10375-10375(1988).
RN [2]
RP SEQUENCE OF 115-214 FROM N.A.
RX MEDLINE=84128872; PubMed=6141822;
RA Pentecost B., Dixon G.H.;
RT "Isolation and partial sequence of bovine cDNA clones for the high-mobility-group protein (HMG-1).";
RL Biosect. Rep. 4:49-57(1984).
RN [3]
RP SEQUENCE OF 1-36.
RX MEDLINE=90306387; PubMed=2365081;
RA Christen T., Bischoff M., Hobi R., Kuenzle C.C.;
RT "High mobility group proteins 1 and 2 bind preferentially to brominated poly(dG-dC).poly(dG-dC) in the Z-DNA conformation but not to other types of Z-DNA."
RL FEBS Lett. 267:139-141(1990).
CC -1- FUNCTION: Binds preferentially single-stranded DNA and unwinds double stranded DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
CC -1- SIMILARITY: Contains 2 HMG box domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X12796; CAA31284.1; -
DR EMBL, M26110; AAA30567.1; -
DR PIR, S01947; S01947.
DR HSSP, P07156; INHN.
DR InterPro, IPR000135; Highmobly_12.
DR InterPro, IPR000910; HMG_12_box.
DR Pfam, PF00505; HMG_box; 2.
DR PRINTS, PR00886; HIGHMOBILITY12.
DR SMART, SM00398; HMG; 2.
DR PROSITE, PS00353; HMG_BOX_1; 1.
DR PROSITE, PS00118; HMG_BOX_2; 2.
KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
FT INIT MET 0
FT DNA_BIND 8 78 HMG_BOX_1.
FT FT 94 162 HMG_BOX_2.
FT DOMAIN 185 214 ASP/GRU-RICH (ACIDIC).
FT CONFLICT 115 119 EHPGL->PGGAV (IN REF. 2).
SQ SEQUENCE 214 AA; 24776 MW; B283A80FC7F0F433 CRC64;

```

Query Match 2.2%; Score 9; DB 1; Length 214;
 Best Local Similarity 100.0%; Pred. No. 0.54;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 293 EDEDEED 301
 |||||
 Db 187 EDEDEED 195

RESULT 7
 HMG1_HUMAN
 ID HMG1_HUMAN STANDARD; PRT; 214 AA.
 AC P09429;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE High mobility group protein 1 (HMG-1).

```

GN HMG1 OR HMG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89160247; PubMed=2922262;
RA Men L., Huang J.K., Johnson B.H., Reek G.R.;
RT "A human placental cDNA clone that encodes nonhistone chromosomal protein HMG-1."
RL Nucleic Acids Res. 17:1197-1214(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96299787; PubMed=8661151;
RA Ferrarri S., Finelli P., Rocchi M., Bianchi M.E.;
RT "The active gene that encodes human high mobility group 1 protein (HMG1) contains introns and maps to chromosome 13."
RL Genomics 35:367-371(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix; and Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggiano N.A., Peters S.J., Abramson R.D., Mollany S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huzyk S.W.,
RA Villalón D.K., Muzny D.K., Sodergren E.J., Lu X., Gibbs R.A.,
RA Rahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Maira W.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 57-64 AND 112-126.
RC TISSUE=breast carcinoma;
RX MEDLINE=97295304; PubMed=9150946;
RA Rasmussen R.K., Ji H., Eddes J.S., Moritz R.L., Reid G.E.,
RA Simpson R.J., Dorow D.S.;
RT "Two-dimensional electrophoretic analysis of human breast carcinoma proteins: mapping of proteins that bind to the SH3 domain of mixed lineage kinase MLK2."
RL Electrophoresis 18:588-598(1997).
CC -1- FUNCTION: Binds preferentially single-stranded DNA and unwinds double stranded DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
CC -1- SIMILARITY: Contains 2 HMG box domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X12597; CAA31110.1; -
DR EMBL, U51677; AAB08987.1; -
DR EMBL, BC003378; AAH03378.1; -
DR EMBL, BC030981; AAH030981.1; -
DR PIR, S02826; S02826.
DR HSSP, P07156; INHN.

```

DR Genew; HGNC:4983; HMG1.
 DR MIM; 163905; -
 DR GO; GO:0000228; C:nuclear chromosome; NAS.
 DR GO; GO:0003697; F:single-stranded DNA binding; NAS.
 DR GO; GO:0006268; P:DNA unwinding; NAS.
 DR InterPro; IPR000135; Hichmobility_12.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 2.
 DR PRINTS; PR00886; HIGHMOBILITY12.
 DR SMART; SM00398; HMG; 2.
 DR PROSITE; PS00353; HMG_BOX_1; 1.
 DR PROSITE; PS50118; HMG_BOX_2; 2.
 DR Nuclear protein; Chromosomal protein; DNA-binding; Repeat.
 KW INIT_MET 0
 FT DNA_BIND 8 78 HMG_BOX_1.
 FT DNA_BIND 94 162 HMG_BOX_2.
 FT DOMAIN 185 214 ASP/GLU-RICH (ACIDIC).
 SQ SEQUENCE 214 AA; 24762 MW; B3CGA80FC7F0FA33 CRC64;
 Query Match 2.2%; Score 9; DB 1; Length 214;
 Best Local Similarity 100.0%; Pred. No. 0.54;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 Db 187 EDEDEDED 195

RESULT 8
 HMG1_PIG STANDARD; PRT; 214 AA.
 ID ID
 AC P12682;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE High mobility group protein 1 (HMG-1).
 GN HMG1 OR HMG1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89050965; PubMed=319113;
 RA Tsuda K.-I., Kikuchi M., Mori K., Waga S., Yoshida M.;
 RT "Primary structure of non-histone protein HMG1 revealed by the
 RT nucleotide sequence."
 RL Biochemistry 27:6159-6163(1988).
 CC -1- FUNCTION: Bands preferentially single-stranded DNA and unwinds
 CC -1- double stranded DNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: Belongs to the HMG1/HMG2 protein family.
 CC -1- SIMILARITY: Contains 2 HMG box domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M21683; AAA31050.1; -.
 CC PIR; A28897; A28897.
 CC HSSP; P07155; 1AAB.
 CC InterPro; IPR000135; Highmobly_12.
 CC InterPro; IPR000910; HMG_12_box.
 CC Pfam; PF00505; HMG_box; 2.
 CC PRINTS; PR00886; HIGHMOBILITY12.
 CC SMART; SM00398; HMG; 2.
 CC PROSITE; PS00353; HMG_BOX_1; 1.
 CC PROSITE; PS50118; HMG_BOX_2; 2.
 KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat.

FT INIT_MET 0 0
 FT DNA_BIND 8 78 HMG_BOX_1.
 FT DNA_BIND 94 162 HMG_BOX_2.
 FT DOMAIN 185 214 ASP/GLU-RICH (ACIDIC).
 SQ SEQUENCE 214 AA; 24785 MW; B29C8A32D8D2C933 CRC64;
 Query Match 2.2%; Score 9; DB 1; Length 214;
 Best Local Similarity 100.0%; Pred. No. 0.54;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 Db 187 EDEDEDED 195

RESULT 9
 KC2B_NEUCR STANDARD; PRT; 333 AA.
 ID ID
 AC Q8TG12;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Casein kinase II beta chain 1 (CK II beta 1).
 GN CKB1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21956574; PubMed=11959847;
 RA Yang Y., Cheng P., Liu Y.;
 RT "Regulation of the Neurospora circadian clock by casein kinase II.";
 RL Genes Dev. 16:994-1006(2002).
 CC -1- FUNCTION: Plays a complex role in regulating the basal catalytic
 CC activity of the alpha subunit (By similarity). A beta 1 subunit and
 CC -1- SUBUNIT: Heterotrimer of 2 alpha subunits, a beta 1 subunit and
 CC a beta 2 subunit.
 CC -1- PTM: Phosphorylated by alpha chain (By similarity).
 CC -1- SIMILARITY: Belongs to the casein kinase 2 beta chain family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF494377; AA014625.1; -.
 CC InterPro; IPR00704; CAS_kinase_II.
 CC Pfam; PF01214; CK II beta; 1.
 CC PRINTS; PR00472; CASNKINASEII.
 CC ProDom; PD003829; CAS_kinase_II; 1.
 CC PROSITE; PS01101; CK2_BETA; 1.
 KW Transferase; Serine/threonine-protein kinase; Phosphorylation.
 SQ SEQUENCE 333 AA; 37107 MW; 1744C2D1F7E1D67C CRC64;
 Query Match 2.2%; Score 9; DB 1; Length 333;
 Best Local Similarity 100.0%; Pred. No. 0.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 Db 68 EDEDEDED 76

RESULT 10
 TRT_DROME STANDARD; PRT; 396 AA.
 ID ID
 AC P19351;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Troponin T, skeletal muscle (upheld protein) (Intended thorax
 DE protein).
 GN UP OR INT.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Asynchronous muscle;
 RX MEDLINE=89141761; PubMed=2852258;
 RA Bullard B., Leonard K., Larkins A., Butcher G., Karlik C.,
 RA Fryberg E.A.;
 RT "Troponin of asynchronous flight muscle.";
 RL J. Mol. Biol. 204:621-637(1988).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=91080155; PubMed=2124273;
 RA Fryberg E.A., Fryberg C.C., Beall C., Saville D.L.;
 RT "Drosophila melanogaster troponin-T mutations engender three distinct
 RT syndromes of myofibrillar abnormalities.";
 RL J. Mol. Biol. 216:657-675(1990).
 CC -!- FUNCTION: Troponin T is the tropomyosin-binding subunit of
 CC troponin, the thin filament regulatory complex which confers
 CC calcium-sensitivity to striated muscle actomyosin ATPase activity.
 CC Mutations in troponin T engender three distinct syndromes of
 CC myofibrillar abnormalities.
 CC -!- SIMILARITY: Belongs to the troponin T family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X54504; CAA838366.1; -;
 DR PIR; S13251; S13251.
 DR Flybase; FBgn0004169; up.
 DR InterPro; IPR001978; Troponin.
 DR Pfam; PF00992; Troponin; 1.
 KW Muscle protein.
 KW DOMAIN
 FT 341 396 ASP/GLU-RICH (HIGHLY ACIDIC).
 SQ SEQUENCE 396 AA; 47333 MW; 3CAAA8AA89493C23 CRC64;
 Query Match 2.2%; Score 9; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 DB 358 EDEDEDED 366
 RESULT 11
 YK18_YEAST
 ID YK18_YEAST STANDARD; PRT; 571 AA.
 AC P36076;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 65.2 kDa protein in MIF2-CYR2 intergenic region.
 GN YK1088W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NX NCBI_TaxID=4912;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Pohl T.M., Pohl F.M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO YEAST AND C.TROPICALIS S182/HAL3. SOME, TO
 CC A.THALIANA HAL3A AND HAL3B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z28088; CAA81926.1; -;
 DR PIR; S37913; S37913.
 DR HSSP; Q9SWE5; IE20.
 DR GerMOnline; I39844; -;
 DR SGD; S0001571; YK1088W.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 KW Hypothetical protein.
 FT DOMAIN 508 570
 SQ SEQUENCE 571 AA; 65238 MW; 9C674C2394EFCBABC64;
 Query Match 2.2%; Score 9; DB 1; Length 571;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 293 EDEDEDED 301
 DB 558 EDEDEDED 566
 RESULT 12
 RGP1_MOUSE
 ID RGP1_MOUSE STANDARD; PRT; 589 AA.
 AC P46061; Q60801;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ran GTPase-activating protein 1.
 GN RANGAP1 OR FUG1.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94148219; PubMed=8314081;
 RA Degregori J., Ruse A., von Melchner H., Rayburn H.,
 RA Priyaranjan P., Jenkins N.A., Copeland N.G., Rulley H.E.;
 RT "A murine homolog of the yeast RNL1 gene is required for
 RT postimplantation development.";
 RL Genes Dev. 8:265-276(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA1B/c;
 RX MEDLINE=95198731; PubMed=7891706;
 RA Ren W., Villamarin A., Shih A., Coutavas E., Moore M.S.,
 RA Locurcio M., Clarke V., Oppenheim J.D., D'Eustachio P., Rush M.G.;
 RT "Separate domains of the Ran GTPase interact with different factors
 RT to regulate nuclear protein import and RNA processing.";
 RL Mol. Cell. Biol. 15:2117-2124(1995).
 CC -!- FUNCTION: GTPase activator for the nuclear Ras-related regulatory
 CC protein Ran, converting it to the putatively inactive GDP-bound
 CC state. Required for postimplantation development.
 CC SUBUNIT: Homodimer. Forms a tight complex in association with
 CC RANBP2 and the ubiquitin-conjugating enzyme E2 (UBC9) (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- PTM: Sumoylated.
 CC -!- SIMILARITY: Contains 6 leucine-rich (LRR) repeats.
 CC -!- SIMILARITY: To fungal RNL1.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U08110; AAA17681.1; -.
CC EMBL; U08857; AAB60517.1; -.
CC PIR; A36983; A36983.
CC PIR; T52070; T52070.
CC PIR; 1K8S; 13-FEB-02.
CC MGD; MGI:103071; Rangapl.
CC InterPro; IPR001611; LRR.
CC InterPro; IPR007091; LRR_FinH.
CC Pfam; PF00560; LRR; 1.
CC GTPase activation; Repeat; Leucine-rich repeat; 3D-structure.
CC REPEAT 48 71 LRR 1.
CC REPEAT 111 134 LRR 2.
CC REPEAT 207 230 LRR 3.
CC REPEAT 235 258 LRR 4.
CC REPEAT 292 315 LRR 5.
CC REPEAT 320 343 LRR 6.
CC DOMAIN 359 399 ASP/GDU-RICH (HIGHLY ACIDIC).
CC CONFLICT 181 LRR R -> A (IN REF. 2).
CC CONFLICT 413 413 S -> L (IN REF. 2).
CC SEQUENCE 589 AA; 63615 MW; 60FAA68B60E9669C CRC64;

Query Match 2.2%; Score 9; DB 1; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.3; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;

QY 293 EDEDEDED 301
Db 378 EDEDEDED 386

RESULT 13
CENB_CRIGR STANDARD; PRT; 606 AA.
ID_CENB_CRIGR STANDARD; PRT; 606 AA.
AC P48988;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
GN CENPB.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Criceulius.
CC NCBI_TaxID=10029;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Ovary;
CC MEDLINE=96254058; PubMed=8652663;
CC Bejarano U.A.; Valdivia M.M.;
CC "Molecular cloning of an intronless gene for the hamster centromere
CC antigen CENP-B."
CC Biochim. Biophys. Acta 1307:21-25(1996).
CC -! FUNCTION: Interacts with centromeric heterochromatin in
CC chromosomes and binds to a specific subset of alphoid satellite
CC DNA, called the CENP-B box. May organize arrays of centromere
CC satellite DNA into a higher order structure which then directs
CC centromere formation and kinetochore assembly in mammalian
CC chromosomes (By similarity).
CC -! SUBUNIT: Homodimer (By similarity).
CC -! SUBCELLULAR LOCATION: Nuclear.
CC -! SIMILARITY: Contains 1 CENPB domain.
CC -! SIMILARITY: Contains 1 CENPB domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC -----
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U20951; AAB06494.1; -.
CC PIR; S70358; S70358.
CC HSSP; P07199; 1BM6.
CC InterPro; IPR004875; CENP-B.
CC InterPro; IPR006895; CENP-B_N.
CC InterPro; IPR006600; CENPB.
CC Pfam; PF04218; CENP-B_N; 1.
CC Pfam; PF03184; DDR; 1.
CC SMART; SM00674; CENPB; 1.
CC Chromosomal protein; Nuclear protein; DNA-binding; Centromere.
CC DNA BIND 1 125 BY SIMILARITY.
CC DOMAIN 404 471 GDU-RICH (ACIDIC).
CC DOMAIN 510 545 ASP/GDU-RICH (ACIDIC).
CC SEQUENCE 606 AA; 66407 MW; 63D0EBE551D48E CRC64;

Query Match 2.2%; Score 9; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.4; Mismatches 0; Gaps 0;
Matches 9; Conservative 0; Indels 0;

QY 293 EDEDEDED 301
Db 528 EDEDEDED 536

RESULT 14
ECM2_HUMAN STANDARD; PRT; 699 AA.
ID_ECM2_HUMAN STANDARD; PRT; 699 AA.
AC O94769;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Extracellular matrix protein 2 precursor (Matrix glycoprotein
DE SCI/ECM2).
GN ECM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC NCBI_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=9909324; PubMed=9790758;
CC Nishiu U.; Tanaka T.; Nakamura Y.;
CC "Identification of a novel gene (ECM2) encoding a putative
CC extracellular matrix protein expressed predominantly in adipose and
CC female-specific tissues and its chromosomal localization to 9q42.3."
CC Genomics 52:378-381(1998).
CC -! SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -! TISSUE SPECIFICITY: Expressed predominantly in adipose tissue as
CC well as female-specific organs such as mammary gland, ovary, and
CC uterus.
CC -! SIMILARITY: Belongs to the small leucine-rich proteoglycan (SLRP)
CC family. Class I subfamily.
CC -! SIMILARITY: Contains 1 WFRC domain.
CC -! SIMILARITY: Contains 12 leucine-rich (LRR) repeats.
CC -! SIMILARITY: Contains 12 leucine-rich (LRR) repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB011792; BAA33958.1; -.
CC Genew: HGNC:3154; ECM2.
CC MIM; 603479; -.
CC GO; GO:0005578; C:extracellular matrix; TAS.

```

DR GO: GO:0005178; F: integrin binding; TAS.
DR GO: GO:0007160; P: cell-matrix adhesion; TAS.
DR InterPro: IPR001611; IIR.
DR InterPro: IPR003591; LRR_LYP.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF00560; LRR; 10.
DR Pfam: PF00093; WVC; 1.
DR PRINTS; PR00019; LEU1RHRPT.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Leucine-rich repeat; Repeat; Glycoprotein; Extracellular matrix;
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 699 EXTRACELLULAR MATRIX PROTEIN 2.
FT DOMAIN 101 158 VWF.
FT REPEAT 335 355 LRR-S 1.
FT REPEAT 356 379 LRR-T 1.
FT REPEAT 382 405 LRR-T 2.
FT REPEAT 406 426 LRR-S 2.
FT REPEAT 427 450 LRR-T 3.
FT REPEAT 451 476 LRR-T 4.
FT REPEAT 477 497 LRR-S 3.
FT REPEAT 524 547 LRR-T 5.
FT REPEAT 596 620 LRR-T 6.
FT REPEAT 621 643 LRR-S 4.
FT REPEAT 644 672 LRR-T 7.
FT REPEAT 673 699 LRR-T 8.
FT DOMAIN 270 281 POLY-GLU.
FT SITE 294 296 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 699 AA; 79789 MW; E4E76A40A5C2742 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 699;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EEEDEDEED 301
DB 283 EEEDEDEED 291

RESULT 15
CH60 PLAFG STANDARD; PRT; 700 AA.
AC P34940;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chaperonin CPN60, mitochondrial precursor.
OS Plasmodium falciparum (isolate FCR-3 / Gambia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5838;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94359534; PubMed=7915821;
RA Hollway S.P., Min W., Inselburg J.I.;
RT "Isolation and characterization of a chaperonin-60 gene of the human
RT malaria parasite Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 64:25-32(1994).
CC -!- FUNCTION: Implicated in mitochondrial protein import and
CC macromolecular assembly. May also prevent misfolding and promote the
CC refolding and proper assembly of unfolded polypeptides generated
CC under stress conditions in the mitochondrial matrix.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X75420; CA53172.1; -.
DR PIR; S38426; S38426.
DR HSP; P06139; 1GRL.
DR InterPro: IPR001844; Chaperonin Cpn60.
DR InterPro: IPR002423; Cpn60/TCF-1.
DR InterPro: IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PROSITE; PS00304; TCOMPLEXTCP1.
KW Chaperone; ATP-binding; Transit peptide; Mitochondrion; Heat shock.
FT TRANSIT 1 9 MITOCHONDRION (POTENTIAL).
FT CHAIN 10 700 CHAPERONIN CPN60.
FT DOMAIN 643 700 ASF/GLU-RICH (HIGHLY ACIDIC).
SQ SEQUENCE 700 AA; 79444 MW; AE1DB362B66D382 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 700;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 293 EEEDEDEED 301
DB 674 EEEDEDEED 682

Search completed: July 22, 2004, 16:54:39
Job time : 15 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:51:37 ; Search time 45 Seconds
(without alignments)
2888.744 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 412
Sequence: 1 MESDETKSISSLISSSSSS.....PLVKKAMDLPNETSVISR 412

Scoring table: OLIGO
Gapco 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10	2.4	546	11	Q61491	Q61491 mus musculu
2	10	2.4	546	11	Q8VENO	Q8VENO mus musculu
3	10	2.4	798	5	O45181	O45181 caenorhabdi
4	10	2.4	6473	5	Q81KH9	Q81KH9 plasmodium
5	9	2.2	163	11	Q8B1C5	Q8B1C5 mus musculu
6	9	2.2	208	11	Q8QY21	Q8QY21 mus musculu
7	9	2.2	211	4	Q9NQJ4	Q9NQJ4 homo sapien
8	9	2.2	215	11	Q88612	Q88612 spalax leuc
9	9	2.2	215	11	Q9QWY6	Q9QWY6 spalax leuc
10	9	2.2	215	11	Q9QX40	Q9QX40 spalax leuc
11	9	2.2	354	11	Q8B1J2	Q8B1J2 mus musculu
12	9	2.2	372	10	Q9S212	Q9S212 arabidopsis
13	9	2.2	382	10	Q38809	Q38809 arabidopsis
14	9	2.2	390	13	Q9DFB9	Q9DFB9 ictalurus p
15	9	2.2	417	4	Q86YR3	Q86YR3 homo sapien
16	9	2.2	435	3	Q06523	Q06523 saccharomyc

17	9	2.2	518	11	Q80W88	Q80W88 mus musculu
18	9	2.2	575	5	Q81205	Q81205 plasmodium
19	9	2.2	589	11	Q91YS2	Q91YS2 mus musculu
20	9	2.2	589	11	Q8C2E3	Q8C2E3 mus musculu
21	9	2.2	589	11	Q7TWM1	Q7TWM1 mus musculu
22	9	2.2	639	4	Q81ZG1	Q81ZG1 homo sapien
23	9	2.2	639	4	Q81UR6	Q81UR6 homo sapien
24	9	2.2	693	4	Q7Z3D0	Q7Z3D0 homo sapien
25	9	2.2	702	5	Q810V3	Q810V3 plasmodium
26	9	2.2	707	11	Q99XK0	Q99XK0 mus musculu
27	9	2.2	774	11	Q8R1Q5	Q8R1Q5 mus musculu
28	9	2.2	774	11	Q8B1I8	Q8B1I8 mus musculu
29	9	2.2	775	11	Q9D6C5	Q9D6C5 mus musculu
30	9	2.2	775	11	Q8BFC7	Q8BFC7 mus musculu
31	9	2.2	792	12	Q9YTL7	Q9YTL7 ateline her
32	9	2.2	1012	16	Q7ULN3	Q7ULN3 rhodoptrell
33	9	2.2	1127	5	Q81S17	Q81S17 dictyostel
34	9	2.2	1404	13	Q8WZV2	Q8WZV2 neurospora
35	9	2.2	1404	13	Q08757	Q08757 gallus gall
36	9	2.2	2075	5	Q81HR5	Q81HR5 plasmodium
37	8	1.9	76	2	P72352	P72352 staphylococ
38	8	1.9	77	16	Q8YF97	Q8YF97 arabidena sp
39	8	1.9	106	11	Q9D822	Q9D822 mus musculu
40	8	1.9	116	11	Q9DAB6	Q9DAB6 mus musculu
41	8	1.9	119	10	Q94190	Q94190 atropa bell
42	8	1.9	159	17	Q9YD24	Q9YD24 aeropyrum p
43	8	1.9	176	2	Q9EZB9	Q9EZB9 staphylococ
44	8	1.9	176	2	Q848T9	Q848T9 enterococc
45	8	1.9	180	2	Q9A103	Q9A103 enterococc

ALIGNMENTS

RESULT 1
Q61491 PRELIMINARY; PRT; 546 AA.
AC Q61491;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE DNA-binding protein.
GN ZEP97.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95137393; PubMed=7835705;
RA Wick M.J., Ann D.K., Lee N.M., Loh H.H.;
RT "Isolation of a cDNA encoding a novel zinc-finger protein from
neuroblastoma x glioma NG108-15 cells.";
RL Gene 152:227-232(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; L20450; AAA65196.1; -;
DR PIR; I49636; I49636.
DR HSSP; P08047; 1SP2.
DR MGD; MGI:105921; Zfp97.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR001909; KRAb.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAb; 1.
DR Pfam; PF00096; Zf-C2H2; 13.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 11.
DR SMART; SM00349; KRAb; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00805; KRAb; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS01571; ZINC_FINGER_C2H2_2; 14.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc; Zinc-finger.

```
SQ SEQUENCE 546 AA; 64148 MW; BEA735BAF5C528B CRC64;
Query Match
Best Local Similarity 2.4%; Score 10; DB 11; Length 546;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 LKRHEITHTK 109
Db 505 LKRHEITHTK 514

RESULT 2
Q8VENO PRELIMINARY; PRT; 546 AA.
AC Q8VENO;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Similar to zinc finger protein 97.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC018101; AAH18101.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB
DR InterPro; IPR007087; ZNF_C2H2.
DR InterPro; IPR007086; ZNF_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; ZNF_C2H2; 10.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZNF_C2H2; 14.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KM Metal-binding: Zinc; Zinc-finger
SQ SEQUENCE 546 AA; 63973 MW; 092E26D554609D5D CRC64;

Query Match
Best Local Similarity 2.4%; Score 10; DB 11; Length 546;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 LKRHEITHTK 109
Db 505 LKRHEITHTK 514

RESULT 3
Q45181 PRELIMINARY; PRT; 798 AA.
AC Q45181;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
```

```
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton B., Hawkins J., Gattung S., Wohlmann P., Elliott G.;
RT "The sequence of C. elegans cosmid K07H8.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047659; AAC04430.1; -.
DR PIR; T33022; T33022.
DR WormPep; K07H8.10; CEI8030.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 1.
KM Hypothetical protein.
SQ SEQUENCE 798 AA; 86819 MW; 6209063CD13FAB26 CRC64;

Query Match
Best Local Similarity 2.4%; Score 10; DB 5; Length 798;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 VEEDEDEED 301
Db 287 VEEDEDEED 296

RESULT 4
Q8IKH9 PRELIMINARY; PRT; 6473 AA.
AC Q8IKH9;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Dynein beta chain, putative.
GN PF14_0626.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.D., Suh B., Peterson J., Anguilo S.,
RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF014826; AAN37239.1; -.
DR GO; GO:0030286; C:dyein complex; IEA.
DR GO; GO:0008567; F:dyein ATPase activity; IEA.
DR GO; GO:0003777; F:microtubule motor activity; IEA.
DR GO; GO:0007018; P:microtubule-based movement; IEA.
DR InterPro; IPR004273; Dynein_heavy.
DR Pfam; PF03028; Dynein_heavy; 1.
SQ SEQUENCE 6473 AA; 771748 MW; BFC848CD45D5D2F8 CRC64;

Query Match
Best Local Similarity 2.4%; Score 10; DB 5; Length 6473;
```

Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 292 VEDEDEDED 301
Db 5443 VEDEDEDED 5452

RESULT 5

Q8BIC5 PRELIMINARY; PRT; 163 AA.
AC Q8BIC5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-OCT-2003 (TREMBLrel. 23, last sequence update)
DE Weakly similar to BA393J16.3.
GN B0114266.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK089532; BAC40916.1; -
DR MGD; MGI:2143362; B0114266.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
SQ SEQUENCE 163 AA; 19461 MW; 1A1AFD8CD74C04 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 163;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 KAFPRKSHL 72
Db 29 KAFPRKSHL 37

RESULT 6

O80YZ1 PRELIMINARY; PRT; 208 AA.
AC O80YZ1;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE B0168F16.1 (Novel protein similar to high-mobility group box 1
DE (Hmgbl1)).
GN B0168F16.1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RC North P., Leaves N., Greystrom J., Coppola M., Manjunath S.,
RA Russell E., Smith M., Strachan G., Tofts C., Boal E., Cobley V.,
RA Hunter G., Kimberley C., Thomas D., Cave-Berry L., Weston P.,
RA Bochevsky M.R.M.;
RL Submitted (FBI-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670999; CAD3017.1; -
DR GO; GO:0000785; C:Chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box_2.
DR PRINTS; PR00886; HIGHMOBILITY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PS00353; HMG_BOX_1; 1.
DR PROSITE; PS0118; HMG_BOX_2; 2.
SQ SEQUENCE 208 AA; 23811 MW; 4DBC87B9516D7B52 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 EEDEDEDED 301
Db 191 EEDEDEDED 199

RESULT 7

O9NQJ4 PRELIMINARY; PRT; 211 AA.
AC O9NQJ4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE D579P20.1 (High-mobility group (Nonhistone chromosomal) protein
DE 1-like 1).
GN HMG111.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RE SEQUENCE FROM N.A.
RC Howden P.;
RL Submitted (FBI-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL160176; CAB92731.1; -
DR HSSP; P07155; 1HMF.
DR GO; GO:000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006377; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box_2.
DR PRINTS; PR00886; HIGHMOBILITY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PS0118; HMG_BOX_2; 2.
SQ SEQUENCE 211 AA; 24238 MW; D2623FDE00FA8355 CRC64;

Query Match 2.2%; Score 9; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 293 EEDEDEDED 301
Db 188 EEDEDEDED 196

RESULT 8

O88612 PRELIMINARY; PRT; 215 AA.
AC O88612;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE High mobility group protein.
GN HMG1.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nanospalax.

```
OX NCBI_TaxID=30637;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K.-L.D., Lum H.-K., Nevo E.;
RT "HMG1 Genes from the Mole Rat Spalax ehrenbergi.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078618; AAC27651.1; -.
DR HSSP; P07156; INHN.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBLTY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PSS0118; HMG_BOX_2; 2.
SQ SEQUENCE 215 AA; 24923 MW; 1C6FB6845CA1E6C8 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 188 EDEDEDEED 196

RESULT 9
ID Q9QWY6 PRELIMINARY; PRT; 215 AA.
AC Q9QWY6;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE High mobility group protein.
GN HMG1.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nannospalax.
OC NCBI_TaxID=30637;
CX [1]
RN SEQUENCE FROM N.A.
RP Lee K.-L.D., Lum H.-K., Nevo E.;
RT "HMG1 Genes from the Mole Rat Spalax ehrenbergi.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078619; AAC27652.1; -.
DR HSSP; P07156; INHN.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBLTY12.
DR SMART; SM00398; HMG; 2.
DR PROSITE; PSS0118; HMG_BOX_2; 2.
SQ SEQUENCE 215 AA; 24822 MW; D20D659274B575B4 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 188 EDEDEDEED 196

RESULT 10
```

```
09QX40
ID Q9QX40 PRELIMINARY; PRT; 215 AA.
AC Q9QX40;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE High mobility group protein.
GN HMG1.
OS Spalax leucodon ehrenbergi (Ehrenberg's mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Nannospalax.
OC NCBI_TaxID=30637;
CX [1]
RN SEQUENCE FROM N.A.
RP Lee K.-L.D., Lum H.-K., Nevo E.;
RT "HMG1 Genes from the Mole Rat Spalax ehrenbergi.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF078820; AAC27653.2; -.
DR HSSP; P07156; INHN.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000135; Highmobility_12.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 2.
DR PRINTS; PR00886; HIGHMOBLTY12.
DR SMART; SM00398; HMG_BOX_1; 1.
DR PROSITE; PSS0118; HMG_BOX_2; 2.
DR PROSITE; PSS0118; HMG_BOX_2; 2.
SQ SEQUENCE 215 AA; 24895 MW; 599FE8A6FDF41F17 CRC64;

Query Match 2.2%; Score 9; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 EDEDEDEED 301
Db 188 EDEDEDEED 196
```

```
RESULT 11
ID Q8BIJ2 PRELIMINARY; PRT; 354 AA.
AC Q8BIJ2;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Weakly similar to BA393J16.3.
GN B114266.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analyses of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK049344; BA033696.1; -.
DR MGD; MGI:2143362; B114266.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 7.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 22, 2004, 16:42:30 ; Search time 54 Seconds

(without alignments)
2155.733 Million cell updates/sec

Title: US-09-831-804-3
2229
Sequence: 1 MSESDETKSISSILSSSSSS.....PLVKAKMDILPNETSVISR 412

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2229	100.0	412	3	AAy93316
2	2225	99.8	412	5	ABP73656
3	445	20.0	564	6	ABU25655
4	445	20.0	564	6	ABJ26255
5	416.5	18.7	1212	4	ABG00399
6	403	18.1	1230	4	AAU30831
7	400	17.9	809	4	AAU38689
8	398	17.9	563	6	ABU96690
9	395.5	17.7	365	3	ABY93317
10	395.5	17.7	409	4	AAU78947
11	395.5	17.7	423	2	AAU91305
12	395	17.7	622	4	AAU78947
13	395	17.7	631	4	AAU78947
14	395	17.7	632	4	ABG18386
15	395	17.7	719	4	ABG16953
16	393.5	17.7	1472	4	AAU31578
17	393	17.6	517	4	ABG03375
18	392.5	17.6	1050	7	ABG09685
19	392.5	17.6	1050	7	ADG32995
20	392	17.6	555	6	ADA54763
21	390.5	17.5	878	4	AAU29528
22	390.5	17.5	1520	4	ABG27130
23	390	17.5	540	7	ADB65594
24	390	17.5	576	6	ADA54797
25	390	17.5	620	6	ABU96725

26	388.5	17.4	501	4	ABB71369
27	388	17.4	568	5	ABR79872
28	387.5	17.4	872	7	ADG37535
29	386	17.3	727	4	ABG16954
30	385.5	17.3	474	6	ABU11782
31	385	17.3	577	5	AAE14680
32	385	17.3	770	4	ABG10091
33	385	17.3	770	4	ABG14920
34	385	17.3	803	4	ABG20106
35	380	17.0	615	6	AAE37047
36	380	17.0	959	4	ABG01956
37	379	17.0	817	4	AAU40475
38	378	17.0	613	4	AAE95862
39	378	17.0	755	4	AAU40916
40	378	17.0	907	4	AAU94428
41	377	16.9	518	6	ABR41469
42	376	16.9	553	7	ADG31307
43	376	16.9	582	6	ABR41407
44	376	16.9	675	5	ABP55423
45	376	16.9	678	7	ADG31847

ALIGNMENTS

RESULT 1
AAy93316
ID AAy93316 standard; protein; 412 AA.

AC AAy93316;
DT 04-SEP-2000 (first entry)

DE A transcription factor designated CATP11A.
KW Transcription factor; CATP11A; DNA-binding protein;
KM ribosomal RNA 5S gene; fungal infection.
XX

OS Candida albicans.

PH Key Location/Qualifiers
FT Misc-difference 193 /note= "Ser encoded by CTG"
FT FT /note= "Ser encoded by CTG"
FT FT /note= "Ser encoded by CTG"
FT FT /note= "Ser encoded by CTG"

PN WO200028037-A1.

PD 18-MAY-2000.

PF 09-NOV-1999; 99WO-FR002739.

PR 10-NOV-1998; 98FR-00014147.

PA (HMRI) HOECHST MARIION ROUSSEL.

PI Bordon-Pallier F, Camier S, Sentenac A;

PT WPI, 2000-376549/32.

PS N-PSDB; AA15398.

XX New nucleic acid encoding Candida albicans transcription factor, useful

XX e.g. in screening for anticyclic agents and for immunization.

XX Claim 12, Page 35-36; 45pp; French.

CC The present sequence represents a Candida albicans transcription factor,
CC designated CATP11A. The polypeptide is a DNA-binding protein, which is
CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as anticyclic agents, to raise an antibody response that is
CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in

CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations
XX
XX
SQ Sequence 412 AA:
Query Match 100.0%; Score 2229; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1e-167;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
Db 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
QY 61 DCDKAFPRKSHLETHIVSHSEKKPFHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
Db 61 DCDKAFPRKSHLETHIVSHSEKKPFHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
QY 121 AFYKHQSIRHHILSVHEKTLTCQCNKVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKRF 180
Db 121 AFYKHQSIRHHILSVHEKTLTCQCNKVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKRF 180
QY 181 QTWSVLQFHIXQSHPKLCPKPCGKGVGKGLSSHMLSHDSTMIKTIWTCDYCVGKFAK 240
Db 181 QTWSVLQFHIXQSHPKLCPKPCGKGVGKGLSSHMLSHDSTMIKTIWTCDYCVGKFAK 240
QY 241 KNEIVEHYNIFHDGNIIPDDLKETEVKKLENLDQSKLANLHLEFTEKLVKVEDEDEE 300
Db 241 KNEIVEHYNIFHDGNIIPDDLKETEVKKLENLDQSKLANLHLEFTEKLVKVEDEDEE 300
QY 301 DSIDEKRSVDVSDMSAQRISIKSFTASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
Db 301 DSIDEKRSVDVSDMSAQRISIKSFTASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
QY 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDDLPLETTSVISR 412
Db 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDDLPLETTSVISR 412
RESULT 2
ABP73656
ID ABP73656 standard; protein; 412 AA.
XX
AC ABP73656;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential protein SEQ ID NO 7493.
XX
XX Fungus; yeast; tetracyclin promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259129P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL,
XX
DR WPI; 2002-566694/60.
DR N-PSDB; AB232206.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.
XX
XX Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.
PS
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a mammalian
CC agent, and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthesis, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Dexamet by
CC the European Patent Office
XX
SQ Sequence 412 AA:
Query Match 99.8%; Score 2225; DB 5; Length 412;
Best Local Similarity 99.8%; Pred. No. 2.1e-167;
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
Db 1 MSESDETKSISSIISSSSSRPKKYYICTYEGCDKAYNRPSELBOHLRTHSNDRPYKCTVD 60
QY 61 DCDKAFPRKSHLETHIVSHSEKKPFHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
Db 61 DCDKAFPRKSHLETHIVSHSEKKPFHCSCGVGKYNRQHLKRHEITHTKSPKCTFENCQE 120
QY 121 AFYKHQSIRHHILSVHEKTLTCQCNKVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKRF 180
Db 121 AFYKHQSIRHHILSVHEKTLTCQCNKVFTRPSKLAQHKLKHGGSPAYQCDHPGCFKRF 180
QY 181 QTWSVLQFHIXQSHPKLCPKPCGKGVGKGLSSHMLSHDSTMIKTIWTCDYCVGKFAK 240
Db 181 QTWSVLQFHIXQSHPKLCPKPCGKGVGKGLSSHMLSHDSTMIKTIWTCDYCVGKFAK 240
QY 241 KNEIVEHYNIFHDGNIIPDDLKETEVKKLENLDQSKLANLHLEFTEKLVKVEDEDEE 300
Db 241 KNEIVEHYNIFHDGNIIPDDLKETEVKKLENLDQSKLANLHLEFTEKLVKVEDEDEE 300
QY 301 DSIDEKRSVDVSDMSAQRISIKSFTASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
Db 301 DSIDEKRSVDVSDMSAQRISIKSFTASLEGSKSVSKLISNGKKINCPKNNCDRMFSREY 360
QY 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDDLPLETTSVISR 412
Db 361 DLRHLLKMHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDDLPLETTSVISR 412
RESULT 3
ABJ25655
ID ABJ25655 standard; protein; 564 AA.
XX
AC ABJ25655;
XX
DT 16-APR-2003 (first entry)
XX


```
XX  Novel human secreted protein #1322.
DE
XX
XX  Human; vaccination; gene therapy; nutritional supplement;
XX  stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX  immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS  Homo sapiens.
XX
XX  WO200179449-A2.
XX
XX  25-OCT-2001.
XX
XX  16-APR-2001; 2001WO-US008656.
XX
XX  18-APR-2000; 2000US-00552929.
XX  26-JAN-2001; 2001US-00770160.
XX
XX  (HYSE-) HYSEQ INC.
XX
XX  Tang YT, Liu C, Drmanac RT;
XX
XX  WPI; 2001-611725/70.
XX
XX  Nucleic acids encoding a range of human polypeptides, useful in genetic
XX  vaccination, testing and therapy.
XX
XX  Claim 20; Page 355; 765pp; English.
XX
XX  The invention relates to novel human secreted polypeptides. The
XX  polypeptides and antibodies to the polypeptides are useful for
XX  determining the presence of or predisposition to a disease associated
XX  with altered levels of polypeptide. The polypeptides are also useful for
XX  identifying agents (agonists and antagonists) that bind to them. Cells
XX  expressing the proteins are useful for identifying a therapeutic agent
XX  for use in treatment of a pathology related to aberrant expression or
XX  physiological interactions of the polypeptide. Vectors comprising the
XX  nucleic acids encoding the polypeptides and cells genetically engineered
XX  to express them are also useful for producing the proteins. The proteins
XX  are useful in genetic vaccination, testing and therapy, and can be used
XX  as nutritional supplements. They may be used to increase stem cell
XX  proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
XX  and/or nerve tissue growth or regeneration; immune suppression and/or
XX  stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
XX  AAU9510-AAU3304 represent the amino acid sequences of novel human
XX  secreted proteins of the invention
XX
XX  Sequence 1230 AA;
XX
XX  Query Match      18.1%; Score 403; DB 4; Length 1230;
XX  Best Local Similarity 32.9%; Pred.No.1.1e-22;
XX  Matches 118; Conservative 45; Mismatches 144; Indels 52; Gaps 17;
XX
XX  20 SRPKYICTYEGCDYAVNRPSTLEQHLFTHSNDRPKYCDVDCDAFFPKSHLETHIVSH 79
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  DB  TREPKYKC--EECGKAFFSQPSHLTHKRNHTEKPKYC--EECGKAFFSQSSTLTTHKTIH 946
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  QY  80 SEKKRPHGVSCKGVNSRQHLKRHBITHT--KSFCTPEMCCGEAFYKHQSL-RHHTLSVH 136
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  DB  947 TGEKPKCECKGKARFKSSTLTTHKTIHTEKPKYC--EECGKARFSQSTLTTHKTIH 1004
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  QY  137 EKTLLCKQCNKVFETPPSLAOKHLKHGGSFAYOCDDHGCFFNPTWSTVLOPHIKQSHPK 196
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  DB  1005 EKPYKCECGKAFNHSKLTTHKTIHTEKPKYC--EECGKAFSSSTLTNGH-KRIHTR 1060
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  QY  197 LKCPKCGKGVCK-----KGLSSHMLSHDSTMTKWTCDYCDVGKFAKKNELVEHNI 250
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  DB  1061 EKPKYC--EGC--GKAFFSQSFTLTGKRKH--TGKPKYKCGEC--GKAFFKSSALTTHKTI 1113
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  QY  251 FHDNIPDILAKTEVKKLNLDDGSKLNLHLETFEKLAVEDEDEDEDESDLDKRSVDV 310
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  DB  1114 IHTGB-----KPKCEKCGKAFNQSSTLTTHKTIHITTPKIH-----T 1151
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY  311 RSDMSAQRSIKSFSTASIGSKSVSKLISNGKKNCPKNNCDMFSERYDLRRHLKMH 369
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  DB  1152 REXPKYKCECGKSNRSTFTPK--HKYIHTGVKLYKC--EECGKFFWSSALTTHKTIH 1206
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX  RESULT 7
XX  AAM38689
XX  ID  AAM38689 standard; protein; 809 AA.
XX
XX  AC  AAM38689;
XX
XX  DT  22-OCT-2001 (first entry)
XX
XX  DE  Human polypeptide SEQ ID NO 1834.
XX
XX  KW  Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX  KW  peripheral nervous system; neuropathy; central nervous system; CNS;
XX  KW  Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX  KW  amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX  KW  chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX  KW  leukaemia.
XX
XX  OS  Homo sapiens.
XX
XX  PN  WO200153312-A1.
XX
XX  PD  26-JUL-2001.
XX
XX  PF  26-DEC-2000; 2000WO-US034263.
XX
XX  PR  23-DEC-1999; 99US-00471275.
XX  21-JAN-2000; 2000US-00488725.
XX  25-APR-2000; 2000US-00552317.
XX  20-JUN-2000; 2000US-00596042.
XX  19-JUL-2000; 2000US-00620312.
XX  03-AUG-2000; 2000US-00653450.
XX  14-SEP-2000; 2000US-00662191.
XX  19-OCT-2000; 2000US-00693036.
XX  29-NOV-2000; 2000US-00727344.
XX
XX  PA  (HYSE-) HYSEQ INC.
XX
XX  PI  Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX  PI  Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX  PI  Zhou P, Goodrich R, Drmanac RT;
XX
XX  DR  WPI; 2001-442253/47.
XX  N-PSDB; AA157845.
XX
XX  PT  Novel nucleic acids and polypeptides, useful for treating disorders such
XX  as central nervous system injuries.
XX
XX  PS  Example 3; SEQ ID NO 1834; 10078pp; English.
XX
XX  The invention relates to human nucleic acids (AA157798-AA161369) and the
XX  encoded polypeptides (AAM38642-AA42213) with nootropic,
XX  immunosuppressant and cyostatic activity. The polynucleotides are useful
XX  in gene therapy. A composition containing a polypeptide or polynucleotide
XX  system, such as peripheral nervous injuries, peripheral neuropathy and
XX  localised neuropathies and central nervous system diseases, such as
XX  Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX  lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX  utilisation of the activities such as: Immune system suppression,
XX  Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX  and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX  assays for receptor activity, arthritis and inflammation, leukaemias and
XX  C.N.S disorders. Note: The sequence data for this patent did not form
XX  part of the printed specification
XX
XX  Sequence 809 AA;
XX
XX  Query Match      17.9%; Score 400; DB 4; Length 809;
```

Best Local Similarity 31.9%; Pred. No. 1.1e-22; Indels 34; Gaps 19;
Matches 121; Conservative 53; Mismatches 171;

QY 2 SESEDEKSISSIISSSS--SRPKKICYEGCDKAYNPISLIEQHILRTHSNDPRKYCTV 59
Db 373 TEGGAFSSSSNLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIH--TGEKPYKCEG- 428

QY 60 DDCDKAFPKRSHLETHIVSHSEKKPHCVGCKGVNSRQHLKHEITHT--KSFKCTPEN 117
Db 429 EECGKAFNPNPSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIHTGKPYKCE--EE 486

QY 118 CGEAFKQHSI--RHHLISVHEKTLTCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPCG 176
Db 487 CGKAFRRSSNLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIH--TGEKPYKCEG- 543

QY 177 FKNFQTSVYLOPH--IKOSHFKLKCPKCGKGVCKGKGLSSHMLSHDSTMIKWTCDYCD 234
Db 544 GKAFNPFSLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIH--TGEKPYKCEG- 599

QY 235 VGR-FAKKNELVHNIFHDGNIPDILLKETEYKLENLDOGSKLNHLETEKLYVE 293
Db 600 -GKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIH--TGEKPYKCEG- 653

QY 294 E-DEDEDESLDEKSDVRSDS--MSAQRSIKSFPTASLEGSKVSXKLISNGKKINCPKN 350
Db 654 ECGKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIH--TGEKPYKCEG- 710

QY 351 NCDRMPREYDLRRHLKMH 369
Db 711 -CGKAFNPNPISLIEQHILRTHSNDPRKYCTV 728

RESULT 8
ABU96690 standard; protein; 563 AA.

AC ABU96690;
XX 25-JUL-2003 (first entry)
DT
XX Human nucleic acid-associated protein (NAAP) #19.
DE
XX Human, nucleic acid-associated protein; cytosolic; antiarteriosclerotic;
KM anticoumarin; neurotropic; neuroprotective; cerebroprotective; anti-HIV;
KM antifibrotic; antiinflammatory; cytotoxic; gene therapy;
KM cell proliferative disorder; cancer; atherosclerosis;
KM neurological disorder; epilepsy; Huntington's disease; stroke;
KM immune disorder; inflammatory disorder; AIDS; allergy;
KM developmental disorder; hypothyroidism; Cushing's syndrome; infection;
KM protein-protein interaction; drug-target interaction;
KM gene expression profile.

OS Homo sapiens.
XX WO2003023003-A2.
XX 20-MAR-2003.
PD
XX 05-SEP-2002; 2002WO-US028540.
PF
XX 07-SEP-2001; 2001US-0317792P.
PR 07-SEP-2001; 2001US-0317912P.
PR 14-SEP-2001; 2001US-032270P.
PR 21-SEP-2001; 2001US-0324040P.
PR 28-SEP-2001; 2001US-0326732P.
PR 19-OCT-2001; 2001US-0346716P.
PR 25-JAN-2002; 2002US-0351749P.
PR 22-FEB-2002; 2002US-0359498P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX Tang YF, Jackson JL, Griffin JA, Elliott VS, Forsythe JF;
PI Becha SD, Richardson TW, Lee EA, Sprague WM, Emerling BM;

PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI Hafalia AU, Sanjanwala B, Margulis JP, Gorvad AE, Lee SY, Ison CH,
PI Raughon MR, Chawla NK, Nguyen DB, Swarnakar A, Zebair-Jadian Y, Shah P,
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang U, Kable AF;
PI Burford N, Ramkumar J;
DR MPI: 2003-313243/30.
XX N-PSDB: ACA98938.
PT New human nucleic acid associated proteome (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
PS Claim 1; Page 255-257; 345pp; English.

XX The invention describes a novel human isolated nucleic acid-associated
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. hypothyroidism, Cushing's syndrome)
CC disorders, or infections. These are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The microarray is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles.
CC This is the amino acid sequence of a novel human nucleic acid-associated
CC protein (NAAP)

Sequence 563 AA:

Query Match 17.9%; Score 398; DB 6; Length 563;
Best Local Similarity 30.2%; Pred. No. 9.3e-23;
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18;

QY 20 SRPKKICYEGCDKAYNPISLIEQHILRTHSNDPRKYCTVDDCKAFPKRSHLETHIVSH 79
Db 196 SGEKPYKCEGKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIH 251

QY 80 SEKPFHCVGCKGVNSRQHLKHEITHT--KSFKCTPENCOEAFYKQSL--RHHLISVH 136
Db 252 TGRKPYKCEGKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIHAG 309

QY 137 EKTLLTCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHGCCKNPTWVLOPH--IKOSH 194
Db 310 EKPYPKCEGKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIHGSE 366

QY 195 PKLNCQPKGCGKGVCKGKGLSSHMLSHDSTMIKWTCDYCDVGFKNKELVEHYNIFHDG 254
Db 367 KPYKCEGKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIHG 421

QY 255 NIPDILLKETEYKLENLDOGSKLNHLETEKLYVEDEDEDESLDEKSDVRSDS 314
Db 422 EKP--YKCEGKAFNQSSTLTTHKKIHTGKPYKCEGKAFNQSSTLTTHKKIHG 467

QY 315 MSAQRST-----KSFASLEGSKVSXKLISNGKKINCPKNCDRMPREYD 361
Db 468 LSKHVIHTGKPYKCEGKAFNQS--SHLTTHMIHTGKPYKCEGKAFNNSI 523

QY 362 LBRHLKMH-----DNLQRTESFINSIEKEE 387
Db 524 LRRHMIHTGKPYKCEGKAFNQS--SHLTTHMIHTGKPYKCEGKAFNNSI 563

RESULT 9
AA93317 standard; protein; 365 AA.
ID AA93317
XX

AC AAY93317;
XX
XX
DT 04-SEP-2000 (first entry)
XX
XX
DE A human transcription factor designated htfIIIA.
XX
XX Human; transcription factor; htfIIIA; DNA-binding protein; transcription;
KW ribosomal RNA 5S gene; transcriptional control; cancer.
XX
OS Homo sapiens.
XX
PN W0200028024-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002738.
XX
PR 10-NOV-1998; 98FR-00014146.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Rocher C;
XX
DR WPI; 2000-387439/33.
XX
DR N-PSDB; AAA15405.
XX
PT New nucleic acid encoding human transcription factor IIA, useful for
PT treatment and diagnosis of cancer and inherited disease.
XX
XX
PS Claim 10; Page 40-41; 49pp; French.
XX
XX
CC The present sequence represents a human transcription factor designated
CC htfIIIA. The polypeptide is probably a DNA-binding protein probably
CC involved in initiating transcription of the gene for ribosomal RNA 5S and
CC maintaining the stability of transcription of other control genes. The
CC htfIIIA polynucleotides and polypeptides are used to make therapeutic or
CC diagnostic compositions for diseases associated with disorders of
CC transcriptional control, particularly cancer or other inherited diseases.
CC The htfIIIA polynucleotide can also be used to detect anomalies in gene
CC transcription, particularly for diagnosis of inherited disease, also for
CC studying diseases involving htfIIIA
XX
SQ Sequence 365 AA;

Query Match 17.7%; Score 395.5; DB 3; Length 365;
Best Local Similarity 31.1%; Pred. No. 8.2e-23;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISSL-----ISSSSSR-----PKYICYEGCDKAYNRPISLEQHRLTHSNDR 53
DB 9 ESVSSLTITADAFIAAGESSAPTPRPALPRFRFCSPDSCSANYSKAWKLDALCKHTGER 68
QY 54 PYKCTVDDCDKAFPRKSHLETHIVSHSEKKEPHGSV--CGKGVNSROHLKXH-ETTH--- 107
DB 69 PFVCDYEGCGKAFIRDYHLSRHLTHGKRFVCAATGCDQKNTKSNLKKHFEKXHENQ 128
QY 108 TKSFKCTPENCOEAFYKQSLR-HHILSVHEKTLTCQ--CNKVFTRPSKLAQHLKXHG 164
DB 129 OKOYICSEFEDCKKTKFKHQQLKIHOCQHTNEPLFKCTOEGCGKHFPASPKLRHAKAHNG 188
QY 165 GSPAYQCHPGCGEKNFQTMVYQEHIKOSHPR----- 196
DB 189 -----YVC-QKGCSPFAKATWTELLKHKVRETHKEEILCEVCRKTKPKRDYLRKQMKTHAER 243
QY 197 --LKCPK--CGKCGVKGKGLSSHMLS-HDDSTMIKITWCDYCDVGK-PAKKNEIVEHNI 250
DB 244 DVCCREBRCGRITVTVENLOSHLSFHEES---RPFVCEHAGCKTKPAMQSLTRH-AV 299
QY 251 FHDGNIPTDLLKETEIVK 268
DB 300 VHD---PDKKKMKLKVKK 314

RESULT 10
AAG75181
ID AAG75181 standard; protein: 409 AA.
XX
XX
AC AAG75181;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5945.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 13.
XX
XX Homo sapiens.
XX
PN W0200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000MO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
XX
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
XX
DR N-PSDB; AAH34586.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX
PS Claim 11; Page 7452-7454; 9803pp; English.
XX
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 409 AA;

Query Match 17.7%; Score 395.5; DB 4; Length 409;
Best Local Similarity 31.1%; Pred. No. 9.6e-23;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISSL-----ISSSSSR-----PKYICYEGCDKAYNRPISLEQHRLTHSNDR 53
DB 53 ESVSSLTITADAFIAAGESSAPTPRPALPRFRFCSPDSCSANYSKAWKLDALCKHTGER 112
QY 54 PYKCTVDDCDKAFPRKSHLETHIVSHSEKKEPHGSV--CGKGVNSROHLKXH-ETTH--- 107
DB 113 PFVCDYEGCGKAFIRDYHLSRHLTHGKRFVCAANCCDQKNTKSNLKKHFEKXHENQ 172
QY 108 TKSFKCTPENCOEAFYKQSLR-HHILSVHEKTLTCQ--CNKVFTRPSKLAQHLKXHG 164
DB 173 OKOYICSEFEDCKKTKFKHQQLKIHOCQHTNEPLFKCTOEGCGKHFPASPKLRHAKAHNG 232

QY 165 GSPAYQCDHGPCRKNQFQTSVLQFHFKQSHPK----- 196
DB 233 ---YVC-QKGSFVAKTWTTELLKHAVERHKEIILCEVCRKTPFRKDYLKQHKTHAPER 287
QY 197 --LKCPK--CGKGCVAKKGLSSHMLS-HDDSTMIKITWTCYCDVGR-FAKKNELVEHYNI 250
DB 288 DVCRCRREGCGRTYTVFVFNQSHILSFHES---RPFVCEHACGKTFPAKQSLTRH-AV 343
QY 251 FHDGNI PDDLKETEYVK 268
DB 344 VHD---PDKKKMKLKVKK 358
RESULT 11
AAR91305
ID AAR91305 standard; protein; 423 AA.
XX AAR91305;
AC
XX
XX 07-JUL-1996 (first entry)
DT
XX
XX
DE Transcription factor-ITIA.
XX
XX Human; transcription factor-ITIA; hTFITIA; DNA binding protein; ribosome;
KM zinc finger; diagnostic; probe; transcription control; antitumour;
KW cancer; therapy.
XX
XX Homo sapiens.
OS
XX
XX EP704526-A1.
PN
XX
XX 03-APR-1996.
PD
XX
XX 05-SEP-1995; 95EP-00113908.
PF
XX
XX 05-SEP-1994; 94JP-00211022.
PR
XX
XX (SANKA) OTSUKA PHARM CO LTD.
PA
XX
XX Fujiwara T, Takeda S, Shimada Y, Ozaki K, Shin S;
PI
XX
XX WPI: 1996-173033/18.
DR N-PSDB; AAT14037, AAT14038.
XX
XX Human Transcription Factor ITI A gene - useful in regulation of
PT transcription and for diagnosis and treatment of e.g. cancer related
PT diseases.
XX
XX Claim 1; Page 8-10; 17pp; English.
XX
XX The sequence represents human transcription factor-ITIA (hTFITIA), a DNA
CC binding protein which is necessary for the initiation of 5S RNA gene
CC transcription, binding to an internal control region of the 5S gene. The
CC protein contains 9 zinc finger domains, which are homologous to the C3H2
CC finger domains of Xenopus TFIIIA, except for the 6th finger domain, which
CC has only 3 amino acid residues between 2 cysteine residues, instead of 5
CC amino acid residues for Xenopus TFIIIA. The protein optionally in
CC recombinant form) and encoding gene may be used in diagnosis,
CC identification or therapy of hereditary diseases such as cancer, or other
CC diseases resulting from abnormal transcriptional control, and to analyse
CC the mechanisms involved in their activity
XX
XX Sequence 423 AA;
SQ
Query Match 17.7%; Score 395.5; DB 2; Length 423;
Best Local Similarity 30.8%; Pred. No. 1e-22;
Matches 98; Conservative 56; Mismatches 95; Indels 69; Gaps 16;
QY 8 KSISGL-----ISSSSSR-----PKKICITVEGCDKAYNPSLLECHLRTHSNDR 53
DB 67 ESVSLITLADARIAAGESSAPTPPPAPPRPFRRCISFPDGSANYSKAMKIDALCKHTGER 126
QY 54 PYKCTVDDCDKAFPRKSHLETHIVGSHSEKPPHCSV--CGKGVNQRQHLKRN-EITW--- 107

DB 127 PFCVDYEGCGKAFIRDYHLSRHILHTHGKPFVCAANGCDQKXNTRKSNLKKEFERKHENO 186
QY 108 TSKFKTEFNQCEAFYKHQSLR-HHLSVHEHTLTQKQ--CNKVFPSPSLAQHKLKHG 164
DB 187 QKQYLSFEDCKKTEFKKHQMKLHQCQNTNEPLFKCTQEGCGKHFPASPSLKRHAAHAG 246
QY 165 GSPAYQCDHGPCRKNQFQTSVLQFHFKQSHPK----- 196
DB 247 ---YVC-QKGSFVAKTWTTELLKHAVERHKEIILCEVCRKTPFRKDYLKQHKTHAPER 301
QY 197 --LKCPK--CGKGCVAKKGLSSHMLS-HDDSTMIKITWTCYCDVGR-FAKKNELVEHYNI 250
DB 302 DVCRCRREGCGRTYTVFVFNQSHILSFHES---RPFVCEHACGKTFPAKQSLTRH-AV 357
QY 251 FHDGNI PDDLKETEYVK 268
DB 358 VHD---PDKKKMKLKVKK 372
RESULT 12
AAM78947
ID AAM78947 standard; protein; 622 AA.
XX AAM78947;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human protein SEQ ID NO 1609.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX
XX W0200157190-A2.
FN
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US004098.
PF
XX
XX 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
XX (HSE-) HXSEQ INC.
PA
XX
XX Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK52080.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
PT
XX
XX Claim 20; Page 3945-3946; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAH80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication

XX
SQ Sequence 622 AA;

Query Match 17.7%; Score 395; DB 4; Length 622;
Best Local Similarity 29.7%; Pred. No. 1.8e-22;
Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;

```
QY 23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDREPYCTVDDCKAFPRKSHLETHIVSHSEK 82
D 201 KPYIC--ECCGKAFKYSALNTWTKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIHTGK 256
QY 83 KPHCSVCGKGVNSROHLKRHEITHT--KSPKCTFENCQAFYKHQSLRHH----- 131
D 257 KPYKCECGKAFNQSSTLTGKKIHTGEKPYKC--ECCGKAFNQSSTLTGKKIHTGEKPY 314
QY 132 -----ILSVH-----EKLITCKQCNKVFTRPSKLAQHLKH----- 162
D 315 YVCECCGKAFKYSRLITTHKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKKHYKCE 374
QY 163 -----HGSPPAYOCDDHPCGCFKNQUTWSVLOPHIKQSH--PKLKCPK 201
D 375 ECGKAFIWSVLTTRHKRVHTGEKPYKCE--CGKAFKYSSTLSH--KRSHTEGKPYKCE 431
QY 202 CGKGVCGKKGILSSHMLSHDSTMIKIWTCDYCDVGK--FAKNELVEHYNI FPHGNIPDL 260
D 432 CGKAFVASSSTLSKHEIHT--TGKKPYKCEC--GKAFNQSSTLTGKKI--HTGEKPY 482
QY 261 LKETEYKYLENLLDQSKLNNLHELET--EKLVEDEDEDEDS--LDEKRSDVRSDSM 315
D 483 YKCECGKAFN--OSSSLTKHKKIHTGEKPYKCECGKAFNQSSTLTGKKIHTGEKPY 539
QY 316 SAQRISKSFASLEGSKSVSKLISNSGSKINCPCKNCDRMFSREYDLRHLKWH 369
D 540 KCECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSHKKIH 589
```

RESULT 13
AAH79931
ID AAM79931 standard; protein; 631 AA.

XX
AC AAM79931;

XX
DT 06-NOV-2001 (first entry)

XX
DE Human protein SEQ ID NO 3577.

XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

XX
PN WO200157190-A2.

XX
PD 09-AUG-2001.

XX
PF 05-FEB-2001; 2001WO-US004098.

XX
PR 03-FEB-2000; 2000US-0046914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.

XX
PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Dimaac RT, Asundi V, Zhou P, Xu C, Cao Y,
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK53064.

XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.

PS Claim 20; Page 388; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, immunomodulatory activity
CC activity, tissue growth factor activity, haematopoiesis regulation
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAH80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication

XX
SQ Sequence 631 AA;

Query Match 17.7%; Score 395; DB 4; Length 631;
Best Local Similarity 29.7%; Pred. No. 1.9e-22;
Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;

```
QY 23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDREPYCTVDDCKAFPRKSHLETHIVSHSEK 82
D 210 KPYIC--ECCGKAFKYSALNTWTKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIHTGK 265
QY 83 KPHCSVCGKGVNSROHLKRHEITHT--KSPKCTFENCQAFYKHQSLRHH----- 131
D 266 KPYKCECGKAFNQSSTLTGKKIHTGEKPYKC--ECCGKAFNQSSTLTGKKIHTGEKPY 323
QY 132 -----ILSVH-----EKLITCKQCNKVFTRPSKLAQHLKH----- 162
D 324 YVCECGKAFKYSRLITTHKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKKHYKCE 383
QY 163 -----HGSPPAYOCDDHPCGCFKNQUTWSVLOPHIKQSH--PKLKCPK 201
D 384 ECGKAFIWSVLTTRHKRVHTGEKPYKCEC--CGKAFKYSSTLSH--KRSHTEGKPYKCE 440
QY 202 CGKGVCGKKGILSSHMLSHDSTMIKIWTCDYCDVGK--FAKNELVEHYNI FPHGNIPDL 260
D 441 CGKAFVASSSTLSKHEIHT--TGKKPYKCEC--GKAFNQSSTLTGKKI--HTGEKPY 491
QY 261 LKETEYKYLENLLDQSKLNNLHELET--EKLVEDEDEDEDS--LDEKRSDVRSDSM 315
D 492 YKCECGKAFN--OSSSLTKHKKIHTGEKPYKCECGKAFNQSSTLTGKKIHTGEKPY 548
QY 316 SAQRISKSFASLEGSKSVSKLISNSGSKINCPCKNCDRMFSREYDLRHLKWH 369
D 549 KCECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSHKKIH 598
```

RESULT 14

ABG18386
ID ABG18386 standard; protein; 632 AA.

XX
AC ABG18386;

XX
DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #18377.

SQ Sequence 719 AA;

Query Match 17.7%; Score 395; DB 4; Length 719;
Best local Similarity 30.0%; Pred. No. 2.2e-22;
Matches 121; Conservative 51; Mismatches 174; Indels 58; Gaps 18;

```
QY      3 ESDETKSISLSSSS--SRPKYICYEGCDKAYNRPSSLLEQHLRTHSNDRPYKCTVD 60
Db      105 ECDKAFKPSLTHKVIHTGEKHYKC--EECGKAFTRSSSLIEHKRSHAGEKPYKC--E 160
QY      61 DCDKAFPRKSHLETHIVSHSEKPPHSCVCGKGVNSROHLKRHEITHTKSPKCFEFCOE 120
Db      161 ECGKAFSKASTLTMAHKTIHAGEKPYKCEECGKAFNRSSNLMHKRIHTGEKPKCECGK 220
QY      121 AFYKHGSL-RHHILSVHEKTLTKQCNKVFTRPSKLAQHKLKHGGSPAYQCD----- 172
Db      221 AFGNFSTLTGKVIHTGEKPYKCECGKAFSWPSSLTEHKRIHAGDKP-YKCECGKTFK 279
QY      173 -----HGG-----CFKRFQTSVLQFH--IKOSHPKLKCPRKCGKGVGKKG 211
Db      280 WSSLTLTGKTIHTGEKPYKCEECGKAFTRSSSLTKHKVIHTGEKHYKCECGKVFSSSS 339
QY      212 ISSHMLSHDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI FHDGNTIPDDLKETEYKLE 270
Db      340 LTHXAIHAGE--KLYKCEEC--GKAFKMSRSLSEHKRI-HTGEKP--YKCECGKAF 390
QY      271 NLLDQSKLNNLHELETEK--LKEVEDEDEEDS--LDEKRSDVRSDSMSAQRSIKSFT 325
Db      391 SWV---SVLNKHKKIHAGKKFYPKCECGKDFNQSSHLTHKRIHTGEKPYKCECGKAF- 446
QY      326 ASLEGSKSVSKLISNSGKINCPKNNCDRMFSREYDLRRLKWH 369
Db      447 -SKASTLTMAHKTIHAGEKPYKC--EECGKAFNRSSNLMHKRIH 487
```

Search completed: July 22, 2004, 16:46:08
Job time : 56 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:44:10 ; Search time 22 Seconds

(without alignments)
1801.405 Million cell updates/sec

Title: US-09-831-804-3

Sequence: 1 MSESDETKSSLSLSSSSSS.....PLYKAMMDLPNETSVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	837.5	37.6	429	2	S20050
2	495	22.2	374	2	T37676
3	400	17.9	803	2	S26823
4	396.5	17.8	595	2	G02075
5	395.5	17.7	363	2	I38937
6	393.5	17.7	1191	2	S35305
7	375.5	16.8	1350	2	S00647
8	374.5	16.8	710	2	I48668
9	373	16.7	542	2	A54661
10	372	16.7	651	2	B32891
11	372	16.7	393	2	JN0533
12	364.5	16.4	428	2	G01496
13	364	16.3	338	2	A32891
14	363.5	16.3	615	2	S06546
15	359.5	16.1	427	2	A35659
16	354	15.9	693	2	I37570
17	353.5	15.9	591	2	S65088
18	351	15.7	474	2	I54338
19	348.5	15.6	439	2	S06556
20	348.5	15.6	707	2	S68858
21	345.5	15.5	335	2	JC1441
22	345.5	15.5	594	2	T12488
23	344	15.4	337	2	S60520
24	344	15.4	580	2	A37107
25	343	15.4	686	2	A34612
26	342	15.3	488	2	S47072
27	341	15.3	336	2	S06578
28	341	15.3	347	2	S00549
29	340	15.3	339	2	JC1442

ALIGNMENTS

```

30      339      15.2      411      2      S10245      finger protein, te
31      338      15.2      247      2      S47070      finger protein HZF
32      338      15.2      383      2      C32891      finger protein 9'
33      337.5      15.1      728      2      A48830      probable transcrip
34      337      15.1      469      2      I38600      zinc finger protei
35      337      15.1      595      2      JC7779      Krueppel-associate
36      336      15.1      553      2      S22954      finger protein zfp
37      335.5      15.1      428      2      S03677      finger protein (cl
38      334.5      15.0      280      2      S06572      finger protein (cl
39      332      14.9      536      2      S06548      finger protein (cl
40      332      14.9      701      2      T14757      hypothetical prote
41      331      14.8      794      2      S59069      zfp protein - mus
42      330.5      14.8      732      2      S47073      finger protein HZF
43      329.5      14.8      496      2      T08674      probable finger pr
44      329      14.8      589      2      I38598      zinc finger protei
45      329      14.8      728      2      A54603      transcription fact

```

RESULT 1

S20050 transcription factor IIA - yeast (Saccharomyces cerevisiae)

N/Alternate names: protein p9677.9; protein YPR186c; TFC2 protein

C/Species: Saccharomyces cerevisiae

C/Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 01-Dec-2000

C/Accession: S20050; A44086; S58816

R/Archambault, J.; Milne, C.A.; Schappert, K.T.; Baum, B.; Friesen, J.D.; Segall, J.

J. Biol. Chem. 267, 3282-3288, 1992

A/Title: The deduced sequence of the transcription factor TFIIA from Saccharomyces ce

A/Reference number: S20050; MUID:92147684; PMID:1737784

A/Accession: S20050

A/Molecule type: DNA

A/Residues: 1429 <ARC>

A/Cross-references: EMBL:M80611; NID:g172902; PID:AB08014.1; PID:g172903

R/Woychik, N.A.; Young, R.A.

Proc. Natl. Acad. Sci. U.S.A. 89, 3999-4003, 1992

A/Title: Genes encoding transcription factor IIA and the RNA polymerase common subuni

A/Reference number: S20050; MUID:92237295; PMID:1570325

A/Accession: A44086

A/Molecule type: DNA

A/Residues: 1429 <MOY>

A/Cross-references: GB:M90638

R/Miller, N.

submitted to the EMBL Data Library, April 1995

A/Description: The sequence of S. cerevisiae cosmid 9677.

A/Reference number: S58816

A/Accession: S58816

A/Molecule type: DNA

A/Residues: 1429 <ML>

A/Cross-references: EMBL:U25841; NID:g786295; PID:g786305; MIPS:YPR186c

C/genetic: SGD:PZFL; TFIIA; TFC2

A/Map position: 16R

C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

C/Keywords: DNA binding; nucleus; transcription factor; zinc finger

Query Match

Best Local Similarity 37.6%; Score 837.5; DB 2; Length 429;

Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;

QY 1 MSESDETKSSLSLIS-SSSSSRPKYICTYGCCKAANRPLLQH-LRTSNDPRYKC 57

DB 23 ISRSSSSLSLNLSTSRSSSSNRPKYFCDDGCKAFRRPSILTEHQLSVHQGLAFCQ 82

QY 58 TVDDCDKAFRRKSHLETIVHSHSEKKPPHSCGKGVNSRQHLKHEHTHTSSPKCTFN 117

DB 83 --DKAKSFYKSHIERHLVTHSDTKPQCSCGCVTTROOLKHEVHTHSFICPERG 140

QY 118 QGAFYKQSLRHHILSVHEKTLTKCCCNKVFTRPSKLQKHHGG--SPAYCCDHPG 175

```

Db      141 CNLRFRKHPOLRAHIIISVHLHKLTCRHCNKSFPQRYLRNHSKHHDPEVENVYQCTPAG 200
QY      176 CFKNFQTSVLOPHIKOSHPEKLCPCGKGVCGKGLSHMLSHDSTMIKWTCDYCDV 235
Db      201 CCEKFRIMSQLOSHIKNDHPKLCPCISKPCVGENGLQMMIITHDLSLVTKMKCHICPD 260
QY      236 GKRAKKNELVEHNINIFH-DQNIIPDDL-LKETEVKYLENLLDQSSKINNHLHELETKLVE 293
Db      261 MFSFRKIDLLTHYGSITHEBDIPLLEIKYKLSDIQQL--VQDHVQLGN-----SKHSNE 312
QY      294 EDEDEEDSDLSDEKRSIDVRSMSAQRISIKFTASLE-GSKSVKGLISNGKLTNCPKNNC 352
Db      313 QDEBKISNRKRRKRLTENNVEFLQNEVDLEKRLGSENGMLNLTVERKXRCPCFNNC 372
QY      353 DRMFSERYDLRRHL--KMHDDNIQLRIEFLNLSIEKEE 387
Db      373 SRFTFKTEKYEKKHIDKAKVHELKLIKILQ-----EKKEE 404

```

RESULT 2

```

T37676
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
C/Accession: T37676
R/McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, October 1999
A/Reference number: Z21736
A/Accession: T37676
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-374 <MCD>
A/Cross-references: EMBL:AL132675; PIDN:CA59689.1; GSPDB:GN00066; SPDB:SPAC144.09C
C/Genetics:
A:Map position: 1
A:Insertions: 50/3
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

```

```

Query Match      22.2%; Score 495; DB 2; Length 374;
Best Local Similarity 32.8%; Pred. No. 5.7e-23;
Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;

QY      15 SSSSSSRPKYITCTYEGCDKAYNRPDLBOHLRTHSNDRPYKCTVDDCKAFPRKSHLET 74
Db      13 SKNLRSAAKIPHCPEYECGKYSRPSLLEOHLRTHSNRPVCDYTGCSKAFYRKSHIKI 72
QY      75 HIVHSEKKPRHGSV--CGKVNRSQHLKREIHTH--TKSPKTPENQGEAFYHQSLRH 130
Db      73 HKRCHTNVPRSPCHYDGDQAFYTOOHLERHIEVRRKRPYACTWEGCDCEGSKHQULRS 132
QY      131 HILSVHEKTL-----TCQCNKVFTRPSKLAQHLKHHGSPAYOCDBPGC--FKNFQTS 184
Db      133 HISACHTHLLYRPTCYDCELRPAFKQLQNHVNRNHEKIIISYSPHESCVGHEGFEKMS 192
QY      185 VLQPHIKOSHPEKLCPCGKGVCGKGLSHMLSHDSTMIKWTCDYCDV---GKPAK 240
Db      193 QLONHIREAHVP--SGSICGRQFKTAHLRHHVVLH--QTTLEERKTYCPMEGCKKSPTR 249
QY      241 KNEIJEHNIFHDGNIIP---DDLKETEVLK--KLEMLLDQSSKINNHLHELETKLVEDE 296
Db      250 SSALKRKHISVHIGMAWPHCDSCGTFGFKHMLQHLNLRGT--CKAAHPYINEGCIKIDG 308
QY      297 EDEEDSDLSDEKRSIDVRSMS--AQRISIKFTASLEGSKSVSKLISNGKLTNCPKNNCDR 354
Db      309 IEGVALIHQKEKELSSNLVSDAKKILNEVVG--HGK-----EAREVSCSPPECNV 358
QY      355 MFSREYDLRRHLKWH 369
Db      359 RFKRLYDMHRHINSH 373

```

RESULT 3

```

S26823
zinc finger protein ZNF43 - human
N/Alternate names: zinc finger protein kox27
C/Species: Homo sapiens (man)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C/Accession: S26823; 137967; S10416
R/Lovering, R.; Trowsdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A/Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell 1.
A/Reference number: S26823; X01D:91279444; PMID:1711675
A/Accession: S26823
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1803 <LOV>
A/Cross-references: EMBL:X59244; NID:G38031; PIDN:CAA11932.1; PID:G38032
R/Thiesen, H.-J.
New Biol. 2, 363-374, 1990
A/Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A/Reference number: 137949; WUID:9145339; PMID:2288909
A/Accession: 137967
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 476-531 <THL>
A/Cross-references: EMBL:X52358; NID:G34160; PIDN:CAA36584.1; PID:G930090
C/Genetics:
A:Gene: GDB:ZNF43; HRP6
A/Cross-references: GDB:128653
A/Map position: 19p13.1-19p12
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C/keywords: DNA binding; zinc finger

```

```

Query Match      17.9%; Score 400; DB 2; Length 803;
Best Local Similarity 31.9%; Pred. No. 7.5e-17;
Matches 121; Conservative 53; Mismatches 171; Indels 34; Gaps 19;

QY      2 SESDETGISLSLSSSS--SRPKYITCTYEGCDKAYNRPDLBOHLRTHSNDRPYKCTV 59
Db      367 TEGCAFARSNSNLTJHKKHITTEKKPYKC--EEGKAFKWSKLTJHKKHITTEKKPYKC-- 422
QY      60 DDCDKAFPRKSHLETTHIVSHSEKKPRHGSVCGKVNRSQHLKREIHTH--KSPKTPEN 117
Db      423 EECGKAFNWSSTLTJHKKHITTEKKPYKC--EEGKAFKWSKLTJHKKHITTEKKPYKC-- 480
QY      118 QGEAFYHQSL--RHIIISVHEKTLTCQCNKVFTRPSKLAQHLKHHGSPAYOCDBPGC 176
Db      481 CGKAFSSSNLTJHKKHITTEKKPYKC--EEGKAFKWSKLTJHKKHITTEKKPYKC--C 537
QY      177 FKNFQTSVLOPH--IKOSHPEKLCPCGKGVCGKGLSHMLSHDSTMIKWTCDYCD 234
Db      538 GKANHPSILTKHRRHITGEKPRYKCEBGKAFQTSNLTJHKKHIT--TGEKPYKCEBG 593
QY      235 VGR-FAKKNELVEHNINIFHDGNIIPDDLKETEVKYLENLLDQSSKINNHLHELETKLVE 293
Db      594 -GKAFQTSNLTJHKKHITTEKKPYKC--YKCEGKAFNQSFTLTJHKKHIT--TEKPYKCE 647
QY      294 E-DEDEEDSDLSDEKRSIDVRSMS--MSAQRISIKFTASLEGSKSVSKLISNGKLTNCPKNN 350
Db      648 ECGKAFKWSSTLTJHKKHITTEKKPYKCEBGKAF--KLSSTLTJHKKHITTEKKPYKCK- 704
QY      351 NCDMFSERYDLRRHLKWH 369
Db      705 -CGKAFNRPNSLTJHKKHIT 722

```

RESULT 4

```

G02075
transcription repressor zinc finger protein 85 - human
C/Species: Homo sapiens (man)
C/Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
C/Accession: G02075
R/Poncellet, D.A.
submitted to the EMBL Data Library, September 1995

```


A:Reference number: G09169
A:Accession: G02075
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-595 <PON>
A:Cross-references: EMBL:U53376; NID:g1017721; PIDN:AAA79179.1; PID:g1017722
C:Genetics:
A:Gene: GDB:ZNR85
A:Cross-references: GDB:132279
A:Map position: 19p12-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 17.8%; Score 396.5; DB 2; Length 595;
Best Local Similarity 31.1%; Pred. No. 8.8e-17;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

QY 10 ISSLSSSSS-SRPKRTCTYEGCDKAYNRPISLEQHLRPHNSNRPKCTVDDCDKAFK 68
DB 186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTGKRIHTEGKPYKC--EECGKAFNQ 241
QY 69 KSHLETHIVSHSEKKPHCSVCGKGVNSROHLKKEHITHT--KSFCTFENCQEAIFYHQ 126
DB 242 SSNLIKHKKIHTGKPYKCECGKAFNRFSTLTGKLIHTEGKPYKC--KECGKAFNRSS 299
QY 127 SL-RHHLISVHEKLTTCQCNKVTRPSKLAQKLIKHGSSPAYQCDHPGCFKNFQTSV 185
DB 300 TLTTHRKIHTEGKPYKCECGKAFNQSSTLTGKIHTGKPYK--YKCKK--CGKAFNQSAAH 356
QY 186 LQFPH--IKQSHPKLCKPCCKGCVGKGLSSHMLSHDSTMTIKITWTDYCDYGFAPKNE 243
DB 357 LTHHEVHTGEEKPYKCKCKGKAFNHFSLTTHKIIH--TGEKPYKCKEC--GKAFNQS 411
QY 244 LVHYNITFDGNIPTDILKETEVKLEMLDQSKLNHLELTELKVEDEDEDEDSL 303
DB 412 TLTGKLIHTEG-----KPYKSECEKAFNQSCL-----TEHKIHTGKPYE--- 455
QY 304 DEKRSVDRSDMSMAQBSIKSFTASL-----EGSK-----SVSKLISNGKINCPKN 350
DB 456 CEKCGKAFNQSSTLTGKRIHTEGKPYKCECGKAFNPSLTLTGKIHTGKPYKC--E 513
QY 351 NCDRMFSREYDLRRHLKWH 369
DB 514 ECGKAFNQSSTLTGKRIH 532

RESULT 5
DNA/RNA-binding protein - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38937
R:Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.
Gene 159, 215-218, 1995
A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIITa.
A:Reference number: I38937; MIMD:95347600; PMID:7622052
A:Accession: I38937
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-363 <REES>
A:Cross-references: EMBL:U02072; NID:g644870; PIDN:AAA75623.1; PID:g644871
C:Superfamily: transcription factor TIIA

Query Match 17.7%; Score 395.5; DB 2; Length 363;
Best Local Similarity 31.1%; Pred. No. 5.9e-17;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISSS-----ISSSSSR-----PKYICTEGCDKAYNRSLLQHLRTSND 53
DB 7 ESSVSLTIDAFIAAESSPAPTPPALPRFICSPDSDANSYKAMKLDHLCRTGGER 66
QY 54 PYKCTVDDCDKAFRRKSHLETHIVSHSEKKPHCSV--CGKGVNSROHLKRIH-ETHT--- 107
DB 67 PFVCDYEGCGKAFIRNHYLSRHILITHTGKPYCANCGANCGDKNTKSNLKKFERKHEHQ 126

QY 108 TKSEKCTFENCQEAIFYHQSLR--HHILSVHEKLTTCQ--CNKVETRPSKLAQKLIKHHG 164
DB 127 QKQYIGCFEDDOCKTFKKHQQLKHQCCHTNEPLFLFKTQEGCGKHFASPSKILKRAKAKHEG 186
QY 165 GSPAYQCDHPGCFKNFQTSVLFHILKQSHPK----- 196
DB 187 ----YVC--QKGCSPVAKTWTELLKRVAFETHEELICVCRKTFKRDYIKQHKMTAPER 241
QY 197 --LKCPK--CGKGVGKGLSSHML--HDSSTMTIKITWTDYCDYK--PAKNEIYHYNIT 250
DB 242 DVCCPREGGGRITVTYFNLIQSHLSFHESS---RPVCEHAGCGKTFAMKQSLTRH-AV 297
QY 251 FHDGNIPTDILKETEVK 268
DB 298 VHD---PDKKMKLKVKK 312

RESULT 6
zinc finger protein ZNF91 - human
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: S35305
R:Beilströmd, E.; Marine, J.C.; Ried, T.; Lecocq, P.; Riviere, M.; Amemiya, C.; Po
EMBO J. 12, 1363-1374, 1993
A:Title: Clustered organization of homologous KRAB zinc finger genes with enhanced exp
A:Reference number: S35305; MIMD:93223677; PMID:8467795
A:Accession: S35305
A:Molecule type: mRNA
A:Residues: 1-191 <BEL>
A:Cross-references: EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186774
A:Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residu
C:Genetics:
A:Gene: GDB:ZNF91; HPF7; HRF10
A:Cross-references: GDB:132284
A:Map position: 19p12-19p12
C:Keywords: DNA binding; zinc finger

Query Match 17.7%; Score 393.5; DB 2; Length 1191;
Best Local Similarity 35.8%; Pred. No. 2.8e-16;
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGCDKAYNRPISLEQHLRPHNSNRPKCTVDDCDKAFRRKSHLETHIVSH 79
DB 905 TREKPYKC--EECGKAFSQSPSHLTGKRIHTEGKPYKC--EECGKAFSQSSTLTGKRIH 960
QY 80 SEKKPHCSVCGKGVNSROHLKKEHITHT--KSFCTFENCQEAIFYHQSL--RHHLISVA 136
DB 961 TGEKPYKCECGKAFNRSSSTLTGKRIHTEGKPYKC--EECGKAFSQSSTLTGKRIH 1018
QY 137 EKLTLQCNKNVTRPSKLAQKLIKHGSSPAYQCDHPGCFKNFQTSVLFHILKQSHPK 196
DB 1019 EKPYKCECGKAFNRSSSTLTGKRIHTEGKPYK--YKCEE--CGKAFISSSTLNGH-KRIHTR 1074
QY 197 ---LKCPKCGKGVGKGLSSHMLSHDSTMTIKITWTDYCDYGFAPKNEIYHYNITFD 253
DB 1075 EKPYKCECGKAFNQSSTLTGKRIH--TGEKPYKCGEC--GKAFKSSALTKKIIHT 1129
QY 254 GNIPDILKETEVKLEMLDQSKLNHLELET 287
DB 1130 GE-----KPYKCEKCKAFNQSSTLTGKRIH 1157

RESULT 7
finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C:Accession: S00647
R:Pauz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A:Title: Xfin: an embryonic gene encoding a multifiingered protein in Xenopus.

A:Reference number: S00647; MUID:88082679; PMID:2826129
A:Accession: S00647
A:Molecule type: mRNA
A:Residues: 1-1350 <RUI>
A:Cross-references: EMBL:X06021
A:Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiator
C:Genetics:
A:Gene: fln
C:Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 375.5; DB 2; Length 1350;
Best Local Similarity 27.3%; Pred. No. 4e-15;
Matches 113; Conservative 65; Mismatches 171; Indels 81; Gaps 20;

QY 11 SLSISSSSSRPKYICTYBGCDAVNPRLLEQHLRTHSNDPRYKCTVDDCDYAFPRKS 70
DB 340 SFLNQQTSHREKPYLCSH--CNKGFIONSDLVGHFRHTGTGERPYQCA--ECHGFTQKS 395
QY 71 HLETHIVSHSKRPFHSCGVNRSQHLKRHEITHT--KSFKCTFENCOEAFYKHQSL 128
DB 396 DLVHKLHTHTGKPKFCHSCDKKFTERSALAKHQRHTHGEKPYKCS--DCKEFTQKSNL 453
QY 129 -RHHLISVHEKTLTCKQCNKVFTRPSKLAQHKLKH-----HGSPAYOCDAHPGCFKN 179
DB 454 ILHQRIHTGERPYKCTLDRTFIQNSDLVKHQVHANPLSDPHANSPHKCSK--CDLT 511
QY 180 FQTVSVLQFHIK--QSHPKLCPKCGKCGVCGKGLSHMLSH-----DD 221
DB 512 FSHWSTFWKSHKSLHSGEKKFCQACCKGKGFQKSDLVKIRIVHTGKPKCLLCKKSPSON 571
QY 222 STMKIM-----TCDCYDVGKFAKKNELVHYNIFHDGNIPDDLKETEYKLENL 272
DB 572 SDLHKHMRHTHGEKPPCYTCD--KSTERSALIKHNR--HTGERPH-----KCSVCKG 623
QY 273 LDQGSKLNNLHELETEKLKVEDEDESDIDEKRSVDYSDMSAQR-----S 320
DB 624 FIQGSAL-----TKSRHTHTGKPYCTQCGKSFIONSDLVKHQRHTHGEKPYHCTEC 676
QY 321 IKSTFASLEGSKSVSKLISNSGKK-INCPKNNCDMFREYDLRHLKWHQ--DNLQRTES 378
DB 677 NKRFTH---EGSLVWHRTHSGEKYRCPQ--CEKTFIQSSDLVHNLVHNGENPPATA 731
QY 379 FLN-SIEKEETPEKGP 393
DB 732 PHEILIRENLTREBP 747

RESULT 8
148668
zinc finger protein 51 - mouse
N:Alternate names: finger protein zfc12
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 08-Dec-2000
C:Accession: I48668; A40984
R:Burke, P.S.; Don, J.; Wolgemuth, D.J.
Mamm. Genome 5, 387-389, 1994
A>Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of C
A:Reference number: I48668; MUID:94319090; PMID:8043957
A:Accession: I48668
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-710 <RES>
A:Cross-references: EMBL:X74855; NID:9488832; PIND:CA52847.1; PID:9488833
R:Crossley, P.H.; Little, P.F.R.
Proc. Natl. Acad. Sci. U.S.A. 88, 7923-7927, 1991
A>Title: A cluster of related zinc finger protein genes is deleted in the mouse embryonic
A:Reference number: A40984; MUID:91376058; PMID:1680234
A:Accession: A40984
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 672-710 <CRO>
A:Cross-references: GB:M74235
C:Genetics:

A:Gene: Zfp-51
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 374.5; DB 2; Length 710;
Best Local Similarity 29.6%; Pred. No. 2.3e-15;
Matches 104; Conservative 54; Mismatches 140; Indels 53; Gaps 15;

QY 25 YICTYBGCDAVNPRLLEQHLRTHSNDPRYKCTVDDCDYAFPRKSLHLETHIVSHSKRP 84
DB 311 YEC--KECGSFTELSTLKKHRYRHTHGEKPKCI--CDSFTTTLTKTHQKHTHTEK 366
QY 85 FHCSCGVNRSQHLKRHEITHT--KSFKCTFENCOEAFKHQSLRHLHS--VHEKTLT 141
DB 367 YKRECKSKSIHSHLRHQRVHTGERPYK--KECKSPHESATLAEHESHTGEKTYK 424
QY 142 CKQCNKVFTRPSKLAQHLKHNGSPAYOCDAHPGCFPNQFQTVSVLQPH--IKQSHPKLC 199
DB 425 CRECDKSFQRAVLRHNRVHTGERPYCKE--CGKSFYCSYLRIHQTHIGKPKYK 482
QY 200 PKCGKCGVCGKGLSHMLSHDSTMIKIMTCDCYDVGKFAKKNELVHYNIFHDGNIPDD 259
DB 483 IECKGSPFNNSYLTTHQKVSGB--KPYRCKECD--KSFYCSYLKAHQSL--HTGEKPYK 537
QY 260 LKETEYKLENLDDQGSKLNNLHELETEKLKVEDEDESDIDEKRSVDYSDMSAQR 319
DB 538 CI-----ECDKSTFYVSHLRTHQRVHTGERPPCTECD-----KSFIRSHLRHQ 583
QY 320 SIKSTFASLEGSKSVSKLISNSGKK-INCPKNNCDMFREYDLRHLKWH 369
DB 584 KI-----HSGEKPYK--RDCDISPSQISLNRHQA 614

RESULT 9
A54661
zinc finger protein ZNF41 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 01-Dec-2000
C:Accession: A54661; I54227
R:Franze, A.; Archidiacono, N.; Rocchi, M.; Marino, M.; Grimaldi, G.
Genomics 9, 728-736, 1991
A>Title: Isolation and expression analysis of a human zinc finger gene (ZNF41) located
A:Reference number: A54661; MUID:91244317; PMID:2037297
A:Accession: A54661
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-542 <FRA>
A:Cross-references: GB:M6548; GB:M92443; NID:9340443; PIND:AAA61312.1; PID:9340444
C:Genetics:
A:Gene: GDB:ZNF41
A:Cross-references: GDB:125865; OMIM:314995
A:Map position: Xp11.3-Xp11.23
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 16.7%; Score 373; DB 2; Length 542;
Best Local Similarity 29.0%; Pred. No. 2.1e-15;
Matches 115; Conservative 59; Mismatches 156; Indels 66; Gaps 21;

QY 14 ISSSSSRPKYICTYBGCDAVNPRLLEQHLRTHSNDPRYKCTVDDCDYAFPRKSLHLE 73
DB 163 IHQKHTHTEKPYEC--NECGKAFTRKSLRMRQRHTHGEKPYVCA--DCKGAFIQKSHFN 218
QY 74 THIVSHSEKRPFHSCGVNRSQHLKRHEITHT--KSFKCTFENCOEAFKHQSLRHLH 131
DB 219 THQRHTHTEKPYECSDCKSTFKSQDLVHQRHTHGEKPYCT--ECKGVTFHNRN----- 272
QY 132 ILVVEKTLT-----CKQCNKVFTRPSKLAQHLKHNGSPAYOCDAHPGCFKNQFQTVSV 185
DB 273 -LTHQKHTHTEKPYKWCACGKAFQDQSNLLKHQHTHGEKPYCN--GCGKAFIWKSR 328
QY 186 LQFHQKSH---PKLCPKCGKCGVCGKGLSHMLSHDSTMIKIMTCDCYDVGK--FAKK 241
DB 329 LKH-OKSHIGRHYECKDCGKAFIQKSTLSVHORIH--TEKPYVCEC--GKAFIQK 382

```
QY 242 MELVAYHYNFHDGNI.PDD-----LlKETBYAKLEENLlDQSKLNNLHE-----284
Db 333 SHFI:AHRI-HHTGEKPEYECSDQCKPFTKSSOL-RVHQXlHTGEPNlCAECGAFITDRSN 440
QY 285 -LETEKLNKEDEDEDEDS-----LDBKESDVSXDSMSAQRST-----KSFTSLSGSKS 333
Db 441 LITHQKlHTREKPYECGDCGKFTFWMSRlNHQKSHTGSRHVECSKCGAFIq-KATLS 498
QY 334 VSKLlSNSGKlNCPKNNCDMEWSREYDlRRHlKMX 369
Db 499 MHQlIHTGKPPAC--TECGKAFITDRSNlIKXlQXKH 532
```

RESULT 10
B32891 finger protein 2, placental - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C:Accession: B32891
R:Belleroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.R.
DNA 8, 377-387, 1989
A:Title: The human genome contains hundreds of genes coding for finger proteins of the x
A:Reference number: A32891; MUID:89377476; PMID:2505992
A:Accession: B32891
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-651 <BET>
A:Cross-references: GB:M27878
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: tandem repeat

C.Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #ext_change 01-Dec-2000
C.Accession: JN0533
R.Brady, J.P.; Platigorsky, J.
Gene 124, 207-214, 1993
A.Title: Cloning and characterization of a novel zinc-finger protein encoding cDNA fro
A.Reference number: JN0533; MUID:93185925; PMID:844344
A.Accession: JN0533
A.Molecule type: mRNA
A1.Residues: 1-393 <BRA>
A2.Cross-References: GB:M98502; NID:g200406; PIDD:AAA9949.1; PID:g200407
A3.Experimental source: eye lens
C.Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C.Keywords: DNA binding; eye lens; zinc finger
F.1-62/Region: acidic
F.63-391/Region: zinc finger

RESULT 11

JN0533
finger protein pMLZ-4 - mouse
C;Species: Mus musculus (house mouse)

A;Accession: U02873; GenBank; U02873.1
A;Map position: 13q12.3-13q13.1
C;Superfamily: transcription factor IIIA

Query Match 16.4%; Score 364.5; DB 2; Length 338;
Best Local Similarity 29.0%; Pred. No. 4.1e-15;
Matches 89; Conservative 49; Mismatches 102; Indels 67; Gaps 11;

C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match	16.1%;	Score 359.5;	DB 2;	Length 427;
Best Local Similarity	32.0%;	Score 411.0;	DB 1;	Length 427;

Local Similarity 32.0%; Pred. No. 1.1e-14;
Matches 99; Conservative 36; Mismatches 121

Matches 99; Conservative 36; Mismatches 121; Indels 53; Gaps 13;

Qy	2	SRPKYICVYEGGDKXVNPRLLEQHLRTHTSNDRPYCVIWDODCKAFPRKSHLETATVSH	79
Dp	132	TVNVPYKC--EAYGRFAPNWSSTLNNKRRILHTGKERPYKC--KECGKAFNQTSHLIRHRIH	187
Qy	80	SEKKDPHCVCVGKGVNSRQHLNKHREITHTKS--FKCTFENCQEAIFYQSTL--PHNLTISVH	136
Dp	188	TEKPYKCECGKAFNOSTLTLTHNLIHGEIPLYKC--EKCVPAPNOSTLTENKHLITHG	245
Qy	137	EKTLNCKOONKVFTRPSKLAONHLKI-----HGSPAY	169
Dp	246	EKRYCEBEGCKAPNBSKLTENHYHTGEKLYKCECGKAFNOSTLTLTKRIHSGEKPY	305
Qy	170	QCDHFGCEPNFQTSVLYQFH--IKOSHPTLKCPKCEGKGVCKGSLSHMLSHDSTMIRI	227
Dp	306	KCEE--CGKAFQFQSNLTDHKKIHTGEKPYKCECGKAFNOSTLNRHVYIH---TGEKP	360
Qy	228	WTCDVDGKFAKKNEIENVNIIFHDGNIIPDLLKETVYKYLENLIDQSKLNNHLELT	287
Dp	361	YKCGEC--GKAFNOSTALNTHKLIHTGENPHKCRBSGKVHNL-----SSKSTCKKIHT	412
Qy	288	-EKL-KVNE	294
Dp	413	GEKLYKCEE	421

Search completed: July 22, 2004, 16:47:55
Job time : 23 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 16:42:55 ; Search time 13 Seconds

(without alignments)
1650.223 Million cell updates/sec

Title: US-09-831-804-3
Sequence: 2229

1 MMSDETKSISLSISSSSS.....PLVKARMDLPNETVISR 412

Scoring table: BIOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	837.5	37.6	429	TF3A_YEAST	P39933 saccharomyc
2	495	22.2	374	TF3A_SCHPO	Q9u15 schizosacch
3	400	17.9	803	ZN42_HUMAN	P17038 homo sapien
4	398	17.9	574	TF3A_HUMAN	Q9p255 homo sapien
5	397.5	17.8	423	TF3A_HUMAN	Q92664 homo sapien
6	396.5	17.8	595	ZN85_HUMAN	Q03923 homo sapien
7	395	17.7	616	ZN93_HUMAN	P35789 homo sapien
8	393.5	17.7	1191	ZN91_HUMAN	Q05481 homo sapien
9	387.5	17.4	839	ZN47_HUMAN	Q96867 homo sapien
10	385	17.3	576	ZN41_HUMAN	Q96832 homo sapien
11	383	17.2	913	ZN28_HUMAN	Q9u1u3 homo sapien
12	376	16.9	570	ZN40_HUMAN	Q9h891 homo sapien
13	376	16.9	751	Z184_HUMAN	Q99676 homo sapien
14	375.5	16.8	1350	XFIN_XENLA	P08045 xenopus lae
15	374	16.8	535	Z257_HUMAN	Q9y2q1 homo sapien
16	373	16.7	470	Z436_HUMAN	Q9p2f3 homo sapien
17	373	16.7	738	ZN64_HUMAN	P51523 homo sapien
18	373	16.7	821	ZN64_HUMAN	P51514 homo sapien
19	372	16.7	393	ZF46_MOUSE	Q09h91 mus musculu
20	367	16.5	818	KR18_MOUSE	Q9h921 homo sapien
21	364	16.3	428	ZN83_HUMAN	P51522 homo sapien
22	363.5	16.3	898	Z071_XENLA	P18751 xenopus lae
23	361	16.2	803	Z226_HUMAN	Q9nyc6 homo sapien
24	360.5	16.2	630	PRD5_HUMAN	Q9nyc1 homo sapien
25	360	16.2	588	Z429_HUMAN	Q86x71 homo sapien
26	359	16.1	1167	Z208_HUMAN	Q43345 homo sapien
27	357.5	16.0	739	YD49_HUMAN	Q9p2f8 homo sapien
28	355	15.9	538	Z155_HUMAN	Q12901 homo sapien
29	354	15.9	498	Z271_HUMAN	Q14591 homo sapien
30	354	15.9	698	Z234_HUMAN	Q14588 homo sapien
31	353	15.8	604	Z305_HUMAN	Q43309 homo sapien
32	351	15.7	474	Z141_HUMAN	Q15928 homo sapien
33	350.5	15.7	783	ZF25_HUMAN	Q9u115 homo sapien

34	350	15.7	670	ZN16_HUMAN	P17020 homo sapien
35	350	15.7	706	ZN25_HUMAN	Q9u1k10 homo sapien
36	349	15.7	489	ZN21_HUMAN	Q9nq28 homo sapien
37	348.5	15.6	439	Z028_XENLA	P18747 xenopus lae
38	348.5	15.6	456	ZN25_HUMAN	Q9u159 homo sapien
39	348.5	15.6	606	ZN14_HUMAN	P16374 mus musculu
40	348.5	15.6	707	ZF60_MOUSE	Q9u1k13 homo sapien
41	348	15.6	617	Z221_HUMAN	Q14709 homo sapien
42	347	15.6	1029	Z197_HUMAN	Q9u1d9 homo sapien
43	346	15.5	623	ZN25_HUMAN	P34695 rana pipien
44	345.5	15.5	335	TF3A_RANPI	P17027 homo sapien
45	345.5	15.5	643	ZN23_HUMAN	

ALIGNMENTS

```
RESULT 1
TF3A_YEAST
ID TF3A_YEAST STANDARD; PRT; 429 AA.
AC P39933;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor TF3A (TF3A).
GN TF3A OR PZF1 OR TF3A1 OR YPR186C OR P9677.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147684; PubMed=1737784;
RA Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA Segall J.
RT "The deduced sequence of the transcription factor TF3A from
RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT TF3A."
RL J. Biol. Chem. 267:3282-3286(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92237295; PubMed=1570325;
RA Woychik N.A., Young R.A.
RT "Genes encoding transcription factor TF3A and the RNA polymerase
RT common subunit RPB6 are divergently transcribed in Saccharomyces
RT cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=6288c / AB972;
RC MEDLINE=97131271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Alberman K., Allen E., Ansoorge W.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunke-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mitrifati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oetner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schater M., Schater M.,
RA Scherens B., Schramm S., Schroeder W., Sclau A.M., Yetteil H.,
RA Uristelar L.A., Ushinsky S., Vierdeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Han J.
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
RL Nature 387:103-105(1997)
-!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR
CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
BINDS THE TRANSCRIBED 5S RNA'S.
```


DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 9.
 KM Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 FW RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 23 47 C2H2-TYPE.
 FT ZN_FING 53 77 C2H2-TYPE.
 FT ZN_FING 83 107 C2H2-TYPE.
 FT ZN_FING 113 138 C2H2-TYPE.
 FT ZN_FING 144 169 C2H2-TYPE.
 FT ZN_FING 204 226 C2H2-TYPE.
 FT ZN_FING 236 261 C2H2-TYPE.
 FT ZN_FING 267 291 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 349 374 C2H2-TYPE.
 SQ SEQUENCE 374 AA; 43851 MW; 7469C701FF08FF6 CRC64;
 Query Match 22.2%; Score 495; DB 1; Length 374;
 Best Local Similarity 32.8%; Pred. No. 1,6e-22;
 Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;

QY 15 SSSSSSRPKYITTYGCDKANKRPSLBQHLTHSNDRPYKCTVDKCDKAFPRKSHLET 74
 DB 13 SKNIRSAKIFHCPCYECCGKYSRPSLBQHLTHSNDRPYKCTVDKCDKAFPRKSHLET 72
 QY 75 HIVESEKPEKPGCV--CGKVNRSQHLKHEITH--TSFKCTFENCQAFYKQSLNH 130
 DB 73 HRCCHTNVAFPSCHYDGCDAQFYTOQLEHIEVHRKPKRYACTWEGCECFKHQQLRS 132
 QY 131 HILSVHEKTL---TKQCNKVFTRPSKLAQHLKHHGSPAYOCDHPCC--FKNPQTS 184
 DB 133 HIASCHTHLPYCTYQDCERFATKQKQNHVNRHAKETISCSHESCVGHEGEXMS 192
 QY 185 VLOFHFKQSHPKKCKCKGKGLSSHMLSHDSTMIKMTQDQDV---GKFAK 240
 DB 193 QLONHIREAHVP--SCSICGRQFKTAHLRHHVTLH--QTTLERKTYHCPMEGCKSKSFR 249
 QY 241 KQELVEHYVIFPDGNP---DULKKEITYK-KLENLDQSGKLNHHELETKAYEER 296
 DB 250 SSALKKHISVIEGNAHFCHDCGTFYGYKMLQRLHREGT-CKAKHKYINECGKHG 308
 QY 297 EDEEDSLDERKSDVSDSMS--AQRISKSFASLEGSKSVSKLISNGKINCPKNNCDR 354
 DB 309 IGVATLHDCKEKELSNLVDAKTIINEVTG--HGK-----EAREYCSPEECNY 358
 QY 355 MSREYDLRRLKWH 369
 DB 359 RFKRLYDMRHLNSH 373

DR ZN43_HUMAN STANDARD; PRT; 803 AA.
 AC P17038; P28160; O96DG1;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein K0X27).
 GN ZNF43 OR ZNF39 OR K0X27.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX MCB1_Taxid:9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=T-cell;
 RX MEDLINE=91279444; Pubmed=1711675;
 RA Lowering R., Trowsdale J.;
 RT "A gene encoding 22 highly related zinc fingers is expressed in
 RL lymphoid cell lines.";
 RN Nucleic Acids Res. 19:2921-2927 (1991).
 RP SEQUENCE FROM N.A.
 RC Tissue=Lymph;

RX MEDLINE=22388257; Pubmed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cassavani T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Tothiyuk S., Carninci P., Prange C.,
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Muliyil S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN (3)
 RP SEQUENCE OF 38-190 FROM N.A.
 RX MEDLINE=91219421; Pubmed=2023909;
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 RA Marciel J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612 (1991).
 RN (4)
 RP SEQUENCE OF 476-531 FROM N.A.
 RC TISSUE=Lymphoid;
 RX MEDLINE=91145339; Pubmed=2288909;
 RA Thiesen H.-J.;
 RT "Multiple genes encoding zinc finger domains are expressed in human T
 RT cells.";
 RL New Biol. 2:363-374 (1990).
 CC -1- FUNCTION: May be involved in transcriptional regulation.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: T AND B CELL LINES.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X59244; CAA41932.1; -;
 DR EMBL; BC006528; AAH06528.1; -;
 DR EMBL; M61869; AAS58674.1; -;
 DR EMBL; X52358; CAA36584.1; -;
 DR PIR; S26823; S26823.
 DR HSSP; P08048; ZNPF.
 DR TRANSFAC; T04986; -;
 DR Genew; HGNC:13109; ZNF43.
 DR MIM; 603972; -;
 DR GO; GO:0003677; F:DNA binding; TAS.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR007087; Znf_C2H2.
 DR InterPro; IPR007086; Znf_C2H2_sub.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; zf-C2H2; 21.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR Prodom; PD000003; Znf_C2H2; 16.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 22.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.

```
CC Eukaryota Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XVII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:143-150(2000).  
CC -1- FUNCTION: May function as a transcription factor.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -1- SIMILARITY: Contains 1 KRAB domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
DR EMBL; AB040906; BAA95997.1; -  
DR HSSP; P07248; IARD.  
DR Genew; HGNC:23707; ZNF492.  
DR InterPro; IPRO01909; KRAB.  
DR InterPro; IPRO07087; ZnF_C2H2.  
DR InterPro; IPRO07086; ZnF_C2H2_sub.  
DR Pfam; PF01352; KRAB; 1.  
DR Pfam; PF00096; ZF-C2H2; 13.  
DR PRINTS; PR00048; ZINCFINGER.  
DR Prodom; PD000003; ZnF_C2H2_2.  
DR SMART; SMO0349; KRAB; 1.  
DR SMART; SMO0355; ZnF_C2H2; 13.  
DR PROSITE; PS50805; KRAB; 1.  
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.  
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 13.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.  
FT NON_TER 1 1  
FT DOMAIN 15 86 KRAB.  
FT FT 184 206 C2H2-TYPE.  
FT ZN_FING 210 234 C2H2-TYPE.  
FT ZN_FING 242 262 C2H2-TYPE.  
FT ZN_FING 268 290 C2H2-TYPE.  
FT ZN_FING 296 318 C2H2-TYPE.  
FT ZN_FING 324 346 C2H2-TYPE.  
FT ZN_FING 352 374 C2H2-TYPE.  
FT ZN_FING 380 402 C2H2-TYPE.  
FT ZN_FING 408 430 C2H2-TYPE.  
FT ZN_FING 436 458 C2H2-TYPE.  
FT ZN_FING 464 486 C2H2-TYPE.  
FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).  
FT ZN_FING 520 542 C2H2-TYPE.  
SQ SEQUENCE 574 AA; 65951 MW; 482E5F1176ECA5B0 CRC64;  
  
Query Match 17.9%; Score 398; DB 1; Length 574;  
Best Local Similarity 30.2%; Pred No.1,3e-16;  
Matches 121; Conservative 48; Mismatches 167; Indels 64; Gaps 18  
  
QY 20 SRPKKYITVYEGCDKAYNRPSLLRCHLTHTSNDRPYKCTVDGDKAFPRKSHLETHIVSH 79  
DB 207 GSEKRYKC--KECGKANVTSTLSLTKRHHTGSKRYKC--EECGKAFFNLISLTLTKIIH 262  
QY 80 SEKKPFHGSGVCAGYNKRQHLEKHITHT--KSFKTEPNCEALYYKOSL-RHHILSIH 136  
DB 263 TGKKRYKCEECCKAPNOSANLTLTKRIHTGEKRYKC--EECGRAFQSSTPLPAHKLIHG 320  
QY 137 EKTLTGCKGNKVFTRPSKLAKGHKLKNHGSSPAYUCDHPGCCFNNGTWSVLGFH--IKOSH 194
```

```

Db      321 EKRYKCECGKAPQSSTLTTHKI -ITGKEFYKCE- CGKAPQSITLTHKRIHSGE 377
QY      195 PKUCPRKCGKGVCKKGLSSHMLSHDSTWIKITWCDYGVGFAKKNELVEHYNI FHDG 254
Db      378 KPYKCECGKAFQSSLTTHKRIHAGE ---KPYKCEVCSKA -FSRSHLTTHKRI HTG 432
QY      255 NIPDDLIKETEYKLEMLLDQSKANLHLELETEKLVBEDEDEDSLDKRSVDYRSDS 314
Db      433 EKP---YKCECGKAPNLSQSLTTHKLIHTGE -KPYKCEBC-----GKAFNQSSST 478
QY      315 MSARST-----KSFYASLEGSKSVSKLISNGSKINCPKNNCDMPFSEYD 361
Db      479 LSHKVIHTGEKPYKYECCGKAFNQ- -SHLTTHKMIHTGEKPYKC- -ECCGKAFNNSI 534
QY      362 LRRRLKXWD-----DNLQRIEPLNISTEKEE 387
Db      535 LNHKMIHTGEKLYKPESCNACDNLAKISKYRNCAGER 574

RESULT 5
TF3A_HUMAN          STANDARD;          PRT;          423 AA.
AC      092664; O12963; Q13097;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Transcription factor IiIA (Factor A) (TFIIIA).
GN      GTF3A.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Fetal brain;
RX      MEDLINE=95309028; PubMed=7789179;
RA      Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,
RT      Fujitawa T., Takahashi E., Shin S., Nakamura Y.
RT      "Molecular cloning, characterization, and chromosomal mapping of a
RT      novel human gene (GTF3A) that is highly homologous to Xenopus
RT      transcription factor IiIA."
RT      Cytogenet. Cell Genet. 70:235-238 (1995).
RN      [2]
RP      SEQUENCE OF 61-423 FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=95347600; PubMed=7622052;
RA      Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.,
RT      Becker K.G.,
RT      "Cloning and expression analysis of a human cDNA homologous to
RT      xenopus TFIIIA."
RT      Gene 159:215-218 (1995).
RN      [3]
RP      CHARACTERIZATION.
RX      MEDLINE=94342241; PubMed=8063702;
RA      Moorefield B., Roeder R.G.,
RT      "Purification and characterization of human transcription factor
RT      IiIA."
RT      J. Biol. Chem. 269:20857-20865 (1994).
CC      -I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC      APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC      CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC      BINDS THE TRANSCRIBED 5S RNA'S. MAY INITIATE TRANSCRIPTION OF THE
CC      5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION
CC      OF OTHER GENES.
CC      -I- SUBCELLULAR LOCATION: Nuclear.
CC      -I- TISSUE SPECIFICITY: Ubiquitous.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial

```

```

CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
Cc      EMBL; D12257; BAA06988.1; -
Cc      DR      EMBL; U120272; AAA75623.1; -
Cc      DR      EMBL; U14134; AAA21873.1; -
Cc      DR      HSSP; P03001; 1TF3.
Cc      DR      TRANSFAC; T04953; -
Cc      DR      Genew; HGNC:4662; GTF3A.
Cc      DR      MIM; 600860; -
DR      GO; GO:0003709; F:RNA polymerase III transcription factor act. .; TAS.
DR      GO; GO:0009303; P:RNA transcription; TAS.
DR      GO; GO:0006383; P:transcription from Pol III promoter; TAS.
DR      InterPro; IPR007087; Znf C2H2.
DR      Pfam; PF00096; zf-C2H2; 9.
DR      SMART; SM00355; Znf C2H2; 9.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR      PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW      Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW      RNA-binding; Repeat; Nuclear protein; Polymorphism.
FT      ZN_FING     98      122
FT      ZN_FING     128     152
FT      ZN_FING     158     183
FT      ZN_FING     190     212
FT      ZN_FING     220     244
FT      ZN_FING     247     271
FT      ZN_FING     275     297
FT      ZN_FING     304     329
FT      ZN_FING     335     359
FT      VARIANT     303     303
FT      CONFLICT    214     214
FT      CONFLICT    221     221
FT      CONFLICT    223     248
FT      CONFLICT    378     387
SQ      SEQUENCE      423 AA; 46847 MW; A627D064A43FB6F0 CRC64;
Query Match      17.8%; Score 397.5; DB 1; Length 423;
Best Local Similarity 31.1%; Pred. No. 9,9e-17;
Matches 99; Conservative 55; Mismatches 95; Indels 69; Gaps 16;
QY      8 KSISSL-----ISSSSSR-----PKYICYEGGDKAYNRPSLEQHLRTHSNDR 53
Db      67 ESVSSTLTADAFIAGESSAPPPRPALPRRTICSPDCCSANYSKAMKLDHLCKRTGR 126
QY      54 PKYCVDDCDKAFRRKSHLETHIVSHSEKRPFCV--CGKGVNSRQHLKRR-ELTH--- 107
Db      127 PFCVCEGCGKAFIDYHLSRILHTGKPFVCAANGCDQKFNKSNLKKHFRKRGHQ 186
QY      108 TSKFKCTFENCOEAFYKHQSLR-HHILSYHEKTLVCKQ--CNKVFTRPSKLAQHLKHNG 164
Db      187 QKQYICSPEDCKCKTKKIQQLKHQCCNNNEPLFKCTQDGCGRHPSFKLKRHAKAHG 246
QY      165 GSPAYQCHPGCGFKNPQTMSVLOFHIKOSHPR----- 196
Db      247 ----YVC-QKGSFPAKWTLELLKAVRETHKEILLCEVCRKTFKRDYLKQMKTHAPER 301
QY      197 --LKCEK--CGKCGVCKKGLSSHMLS-HDSTWIKITWCDYGVG-FKKKELVEHYNI 250
Db      302 DVCRCPRCGRTTYTNNLSHILSFHEES--BPVCEHAGCGKTFAMKOSLTRH-AV 357
QY      251 FHDGNIIPDILKETEYKK 268
Db      358 VHD---PDKKKKLVKK 372

RESULT 6
ZN85_HUMAN
AC      003923;
ID      ZN85_HUMAN          STANDARD;          PRT;          595 AA.
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Zinc finger protein 85 (Zinc finger protein HPF4) (HTEFL).
GN ZNF85.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Placenta;
RA MEDLINE=99053537; PubMed=9839802;
RX Poncelet D.A., Bellefroid E.J., Bastiaens P.V., Demotie M.A.,
RA Marne J.C., Penderville H., Alami Y., Devos N., Lecocq P.J., Ogawa T.,
RT Muller M., Marital J.A.;
RT "functional analysis of ZNF85 KRAB zinc finger protein, a member of
RT the highly homologous ZNF91 family.";
RL DNA Cell Biol. 17:931-943(1998).
RN [2]
RP SEQUENCE OF 1-196 FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Marital J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -1- FUNCTION: Transcriptional repressor.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed preferentially in testicular
CC tissues.
CC -1- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U35376; AAA79179.1; -;
DR EMBL: M61866; AAA52689.1; -;
DR EMBL: M61868; AAA56671.1; -;
DR PIR: C39384; C39384.
DR PIR: G02075; G02075.
DR HSSP: P08048; ZNF.
DR TRNSPAC: T04990; -;
DR Genew: HGNC:13160; ZNF85.
DR MIT: 603899; -;
DR GO: GO:0005634; C:nucleus; TAS.
DR GO: GO:0003714; F:transcription co-repressor activity; TAS.
DR GO: GO:0003700; F:transcription factor activity; TAS.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR007087; ZnF_C2H2.
DR InterPro: IPR007086; ZnF_C2H2_sub.
DR Pfam: PF01352; KRAB. 1.
DR Pfam: PF00096; ZF-C2H2. 15.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; ZnF_C2H2. 13.
DR SMART: SM00349; KRAB. 1.
DR SMART: SM00355; ZnF_C2H2. 15.
DR PROSITE: PS50805; KRAB. 1.
DR PROSITE: PS50028; ZINC_FINGER_C2H2_1. 14.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2. 15.
DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat; Repressor.
FT DOMAIN 4 75 KRAB
FT ZN_FING 146 168 C2H2-TYPE (DEGENERATE).
FT ZN_FING 174 196 C2H2-TYPE.
FT ZN_FING 202 224 C2H2-TYPE.
FT ZN_FING 230 252 C2H2-TYPE.
FT ZN_FING 258 280 C2H2-TYPE.

FT ZN_FING 286 308 C2H2-TYPE.
FT ZN_FING 314 336 C2H2-TYPE.
FT ZN_FING 342 364 C2H2-TYPE.
FT ZN_FING 370 392 C2H2-TYPE.
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 426 448 C2H2-TYPE (DEGENERATE).
FT ZN_FING 454 476 C2H2-TYPE.
FT ZN_FING 482 504 C2H2-TYPE.
FT ZN_FING 510 532 C2H2-TYPE.
FT ZN_FING 538 560 C2H2-TYPE.
FT ZN_FING 566 588 C2H2-TYPE.
FT CONFLICT 84 84 R -> Q (IN REF. 2; AAA52689).
FT CONFLICT 115 115 R -> I (IN REF. 2; AAA58671).
FT CONFLICT 177 177 T -> R (IN REF. 2; AAA58671).
FT CONFLICT 184 184 G -> R (IN REF. 2; AAA58671).
SQ SEQUENCE 595 AA; 68718 MW; 44AA0A236D2D43B CRC64;
Query Match 17.8%; Score 396.5; DB 1; Length 595;
Best Local Similarity 31.7%; Pred. No. 1.6e-16;
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;
QY 10 ISSLSISS--SSRPKYICTYEGCDKAYNRPSTLLEQHTHNSMDRPYKCTVDDODKAFPR 68
DB 186 ISCLTEHSRIHTVNFYKC--EECGKAFWNSSTLYHKRIHTGEKPYKC--EECGKAFNQ 241
QY 69 KSHLETHIVSHSEKPKFPCVCGVNSROHLKRHEITH--KSFCTPENCOPAFYHQ 126
DB 242 SSNLIHKHKTHTGEKPYKCECGKAFRPSFTLYHKRIHTGEKPYKC--KECGKAFRRSS 299
QY 127 SL-RHHLIVSHEKTLTCKQKNVFTPRSKLAQHKLKHGSSPAYQCDHPGCFKNFQTVSV 185
DB 300 TLFTHRKIHTGEKPYKCECGKAFKQSSNLTTHKIHTGKRP--YKCKK--GKAFNQSAAH 356
QY 186 LGHH--IKQSHPKLCKPGKCGVGGKGLSHLSHSDSMIKITWCDYDVGFAKNE 243
DB 357 LTTHEVHTGEKPYKCECGKAFNHSHTLTHKTIH--TGEKPYKCEK--GKAFHSS 411
QY 244 LVEHYNIPTHGNIPTDILKEVEVKLEMLDQSKLNNLHELETKEKLVDEDEDEDSLT 303
DB 412 TLTKHKTIHTGE-----KPKSKCECGKAFNQSXL-----TEHKKIHTGEKPYE--- 455
QY 304 DEKRSVRSWSQSAQSKISFTFASL-----EGSK-----SVSKLSNSGKLNCPKN 350
DB 456 CEKCGKAFNQSNNLTRHKSHTEKPYKCECGKAFWPSFTLYHKRIHTGEKPYKC--E 513
QY 351 NCDRMFSREYDIAHKLKMH 369
DB 514 ECGKAFNQSRLTKHKTIH 532
RESULT 7
ZN93_HUMAN STANDARD; PRT; 616 AA.
ID_ZN93_HUMAN
AC P35789; Q9Y2N8;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 93 (Zinc finger protein HRF34) (Fragment).
GN ZNF93.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Sakaidais G.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Scott D., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Dangnan L.,
RA Ertler A., Christensen M., Georgescu A., Avila J., Attix C.,
RA Andreise T., Trankheim M., Amico-Keller G., Coefield J., Duarte S.,
RA Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA Carrano A.V.;

RT "Sequence analysis of a 5.7 Mb region in 19p13.1 between OLF and
 RT DJ98455." ;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-191 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncet D.A., Lecocq P.J., Revelant O.,
 RA Martial J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -1- FUNCTION: May be involved in transcriptional regulation.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AC007204; AAD2281.1; .
 DR EMBL: M61873; AA83548.1; .
 DR HSSP: P08046; 1A1H.
 DR Genew: HGNC:13169; ZNF93.
 DR MIM: 603975; .
 DR GO: GO:0005634; C:nucleus; NAS.
 DR GO: GO:0003700; F:transcription factor activity; NAS.
 DR GO: GO:0008270; F:zinc ion binding; NAS.
 DR GO: GO:0006355; P:regulation of transcription; NAS.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR007087; Znf_C2H2.
 DR InterPro: IPR007086; Znf_C2H2_sub.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf-C2H2; 16.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR PRODom: PD000003; Znf_C2H2; 7.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; ZNF_C2H2; 16.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 16.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 17.
 KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
 KW Nuclear protein; Repeat.
 FT NON_TER 1
 FT DOMAIN 1
 FT 1 71 KRAB.
 FT 163 C2H2-TYPE (DEGENERATE).
 FT 169 191
 FT 197 219 C2H2-TYPE.
 FT 225 247 C2H2-TYPE.
 FT 253 275 C2H2-TYPE.
 FT 281 303 C2H2-TYPE.
 FT 309 331 C2H2-TYPE.
 FT 337 359 C2H2-TYPE.
 FT 365 387 C2H2-TYPE.
 FT 393 415 C2H2-TYPE.
 FT 421 443 C2H2-TYPE.
 FT 449 471 C2H2-TYPE.
 FT 477 499 C2H2-TYPE.
 FT 505 527 C2H2-TYPE.
 FT 533 555 C2H2-TYPE.
 FT 561 583 C2H2-TYPE.
 FT 589 611 C2H2-TYPE.
 FT 611 71 S -> SGP (IN REF. 2).
 FT 71 85 Q -> H (IN REF. 2).
 FT 85 115 S -> R (IN REF. 2).
 FT 115 132 C -> S (IN REF. 2).
 FT 132 149 V -> D (IN REF. 2).
 FT CONFLICT 149

SQ SEQUENCE 616 AA; 70572 MW; DF30983AB61160A CRC64;
 Query Match 17.7%; Score 395; DB 1; Length 616;
 Best Local Similarity 29.7%; Pred. No. 2,1e-16;
 Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;
 QY 23 KKLYCTYEGCDKAYNRPSTLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSSEK 82
 DB 195 KPYIC--EEGCKAFKYSALNTHKRIHTGKPKYC--DKDKAFIASSTLSKRIHTGK 250
 QY 83 KPHCVCGKGVNSROHLKRHEITHT--KSFCTFENCQAFYKHOSLRHH----- 131
 DB 251 KPYKCECGKAFNOSTLTGKKIHTGKPKYC--EEGKAFNOSTLTGKKIHTGKPK 308
 QY 132 -----ILSVH-----EKTLCKCNVFPSTGLAQHKYH----- 162
 DB 309 YVCEBGCKAFKYSRLITLTKRIHTGKPKYCKNGKAFIASSTLSHHEFTIMGKPKYCE 368
 QY 163 -----HGSPAYQCDHPGCFKPFQTSVLQFHLKOSH--PKLKCPK 201
 DB 369 ECGKAFIWSVLTNRKRVHTGKPKYCEE--CGKAFKYSSTLSH--KRSHTGKPKYCEE 425
 QY 202 CGKCGYCKGKGLSNNLSHDDSTMIKIWTCDYCDVK--PAKKMELVHYNIFFHDGNTPDL 260
 DB 426 CGKAFVASSSTLSKHEIHT--TGKPKYKCEBC--GKAFNOSTLTGKKI--HTGKPK--- 476
 QY 261 LKETEYKTKLENLDDQSKNNLHELET--EKLVKVEDDEDEBDS--LDKESDVSDSM 315
 DB 477 YKCEBGCKAFN--QSSSLTKHKIHTGKPKYCEGKAFNOSTLTGKKIHTGKPKY 533
 QY 316 SAQRISKSFYASLEGSKYSKLIISNGKCKINGCKNNCDMPFSREYDLRHLKWH 369
 DB 534 KCEBGCKAF--HLSTHLTTHKLIHTGKPKYRC--RCGKAFNHSATLSHKLIH 583
 RESULT 8
 ID ZN91 HUMAN STANDARD; PRT; 1191 AA.
 AC 005481;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 91 (Zinc finger protein HTF10) (HPR7).
 GN ZNF91.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93223677; PubMed=8467795;
 RA Bellefroid E.J., Marine J.C., Ried T., Lecocq P.J., Riviere M.,
 RA Ametia C.T., Poncet D.A., Coulle P.G., de Jong P.J.;
 RA Szpirer C., Ward D.C., Martial J.A.;
 RT "Clustered organization of homologous KRAB zinc-finger genes with
 RT enhanced expression in human T lymphoid cells." ;
 RL EMBO J. 12:1363-1374(1993).
 RN [2]
 RP SEQUENCE OF 15-204 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncet D.A., Lecocq P.J., Revelant O.,
 RA Martial J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 1 KRAB domain.
 CC -1- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS
 CC DERIVED FROM THE TRANSLATION OF AN ALU REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; L11672; AAA59469.1; -.
DR EMBL; M61871; AAA58672.1; ALT_SEQ.
DR PIR; S35305; S35305.
DR HSSP; P08047; ISP2.
DR GeneW; HGNC:13166; ZNF91.
DR MIM; 603973; -.
DR GO; GO:0003634; C:nucleus; NAS.
DR GO; GO:0003700; F:transcription factor activity; NAS.
DR GO; GO:0008270; F:zinc ion binding; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf_C2H2; 34.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 20.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 33.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 31.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 35.
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat.

FT DOMAIN 13
FT ZN_FING 134
FT ZN_FING 176
FT ZN_FING 179
FT ZN_FING 200
FT ZN_FING 208
FT ZN_FING 232
FT ZN_FING 238
FT ZN_FING 260
FT ZN_FING 266
FT ZN_FING 288
FT ZN_FING 294
FT ZN_FING 316
FT ZN_FING 322
FT ZN_FING 344
FT ZN_FING 350
FT ZN_FING 372
FT ZN_FING 378
FT ZN_FING 400
FT ZN_FING 406
FT ZN_FING 428
FT ZN_FING 434
FT ZN_FING 456
FT ZN_FING 462
FT ZN_FING 490
FT ZN_FING 512
FT ZN_FING 518
FT ZN_FING 540
FT ZN_FING 546
FT ZN_FING 568
FT ZN_FING 574
FT ZN_FING 596
FT ZN_FING 602
FT ZN_FING 624
FT ZN_FING 630
FT ZN_FING 652
FT ZN_FING 658
FT ZN_FING 680
FT ZN_FING 686
FT ZN_FING 708
FT ZN_FING 714
FT ZN_FING 736
FT ZN_FING 742
FT ZN_FING 770
FT ZN_FING 792
FT ZN_FING 798
FT ZN_FING 820
FT ZN_FING 826
FT ZN_FING 848
FT ZN_FING 854
FT ZN_FING 876
FT ZN_FING 904
FT ZN_FING 910
FT ZN_FING 932
FT ZN_FING 938
FT ZN_FING 960
FT ZN_FING 966
FT ZN_FING 988
FT ZN_FING 994
FT ZN_FING 1022
FT ZN_FING 1044
FT ZN_FING 1050
FT ZN_FING 1078
FT ZN_FING 1100
FT ZN_FING 1128
FT ZN_FING 1134
FT ZN_FING 1156
SQ SEQUENCE 1191 AA; 137225 MW; 581056B811B8716D CRC64;

Query Match 17.7%; Score 393.5; DB 1; Length 1191;
Best Local Similarity 35.8%; Pred. No. 5,2e-16;
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGGDAXNRPSSLLEQHLRTHSNDRPYKCTVDDCKAFPRKSHLETHIVSH 79
Db 905 IREKPYKC--EECGKAFSPQSHLTTRKHTGEPKVC--EECGKAFSQSSTLTTRKILH 960
QY 80 SEKKPHCGVCGKGVNSRQHLKRHEITHT--KSKCTFENCQAFYKHQSL-RHILSYA 136
Db 961 TGEKPYKCECGKAFKSSSTLTTRKILHTGEPKVC--EECGKAFSQSSTLTTRHTMTG 1018
QY 137 EKLTLTKQCNKYPTRPSKLAQHLKHHGSPAQCCHPGCFKNPQTWSVLQFIHQSHPK 196
Db 1019 EKPVKCECGKAFNRSSKLTTRKILHTGEPKVC--CGKAFISSSTLTNGH-KRILHR 1074
QY 197 ---LKCPKCGKGVCGKKGKLSHMLSHDSTMIKIVTCYCDVGKFAKKNELVENVIFHD 253
Db 1075 EKPYKCECGKAFSQSSTLTTRHGRHL---TGEKPYKGC--GXAFRESSALTTRKILHT 1129
QY 254 GNIPDDLKRETEYKLENLIDQSKLNINHELET 287
Db 1130 GE-----KPYKCECGKAFNQSSILTTRHKIHT 1157

RESULT 9
ID 2347 HUMAN STANDARD; PRT; 839 AA.
AC 096SER; 08TCN1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 347 (Zinc finger 111).
GN ZNF347 OR ZNF111.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Aitken C.J., Nicholson G.C.;
RT "RANKL regulated zinc finger protein in osteoclastogenesis.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 305-839 FROM N.A.
RC TISSUE=Brain;
RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May function as a transcription factor.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: Contains 1 KRAB domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AY029765; AAA37403.1; -.
DR EMBL; AL713691; CAD28491.1; -.
DR GeneW; HGNC:16447; ZNF347.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf_C2H2; 18.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 19.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.

DR PROSITE; PS50157; ZINC FINGER C2H2 2; 20.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 8
 FT ZN_FING 261 283 KRAB.
 FT ZN_FING 289 311 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 317 339 C2H2-TYPE.
 FT ZN_FING 345 367 C2H2-TYPE.
 FT ZN_FING 373 395 C2H2-TYPE.
 FT ZN_FING 401 423 C2H2-TYPE.
 FT ZN_FING 429 451 C2H2-TYPE.
 FT ZN_FING 457 479 C2H2-TYPE.
 FT ZN_FING 485 507 C2H2-TYPE.
 FT ZN_FING 513 535 C2H2-TYPE.
 FT ZN_FING 541 563 C2H2-TYPE.
 FT ZN_FING 569 591 C2H2-TYPE.
 FT ZN_FING 597 619 C2H2-TYPE.
 FT ZN_FING 625 647 C2H2-TYPE.
 FT ZN_FING 653 675 C2H2-TYPE.
 FT ZN_FING 681 703 C2H2-TYPE.
 FT ZN_FING 709 731 C2H2-TYPE.
 FT ZN_FING 737 759 C2H2-TYPE.
 FT ZN_FING 765 787 C2H2-TYPE.
 SQ SEQUENCE 839 AA; 95771 MW; F7F7C07069E7844E CRC64;
 Query Match 17.4%; Score 387.5; DB 1; Length 839;
 Best Local Similarity 30.7%; Pred. No. 8e-16;
 Matches 112; Conservative 46; Mismatches 154; Indels 53; Gaps 16;
 QY 11 SSI-ISSSSSRPRKTYCTEGGCDKAVNRPDLLEQHLRTSHNDRPYKCTWDDCKAFPRK 69
 DB 386 SSIAIHQATHSGKPKYC--NECGKVFQNSHLTNHRIHTGKPKYC--NECGKAFGVR 441
 QY 70 SHLETVIVSHSEKKPHFSCVCGVNSRHLKREHITHT--KSFCTFENCQAFYKHOS 127
 DB 442 SSIAIHQATHSGKPKYC--NECGKVFQNSHLTNHRIHTGKPKYC--NECGKAFGVR 499
 QY 128 L-RHHILSVHEKLTGCKQCKVTRSPSKLAQKLTAKHGGSPAYQCHPGCFKNPQTWVL 186
 DB 500 LTHQVHTGKPKYCNECGKVFQNSHLTNHRIHTGKPKYC--NECGKAFGVR 556
 QY 187 OFH--TKQSHPKLKPCGCGVCGKKGKSSHMLSHDSTIMIKWTGDDYDVGFAKKNEL 244
 DB 557 TTHQVHTGKPKYCNECGKVFQNSHLTNHRIHTGKPKYC--NECGKAFGVR 611
 QY 245 VEHYNIETHDGNIPDDLKETEYKLEMLDQSGSKLNNLHELETEKLVEDDEDEBDSID 304
 DB 612 LSRHQRITHTGKPKYC--YKTYNEGK--AFSEHSLNLTTHQVHTGKPKYCNE----- 657
 QY 305 EKRSDVRSDSMAQKSIKFTSLBSKSVSKLISNGSKKINCPKXNCDRMFSREYDLRR 364
 DB 658 -----CGKVFQNSHLTNHRIHTGKPKYCNE----- 698
 QY 365 HLKWH 369
 DB 699 HQRVH 703
 RESULT 10
 2431 HUMAN STANDARD; PRT; 576 AA.
 ID 2431 HUMAN
 AC O8TF32; O8TFW4; (Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DB Zinc finger protein 431.
 GN ZNF431 OR KIAI1969.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Placentalia; Eumarchia; Homiidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
 RA MEDLINE=21842142; Pubmed=11853319;
 RX Nagase T., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XXII.
 RT The complete sequences of 50 new cDNA clones which code for large
 RT proteins.";
 RL DNA Res. 8:319-327(2001).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; Pubmed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gutarra P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maita M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RA and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May be involved in transcriptional regulation.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPFL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: Contains 13 C2H2-type zinc fingers.
 CC -1- SIMILARITY: Contains 13 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL; AB07849; BAB8555.1; ALT INIT.
 DR EMBL; BC040506; AAH40506.1; --
 DR Genbank; HGNC:20809; ZNF431.
 DR InterPro; IPR001909; KRAB.
 DR InterPro; IPR007087; Znf_C2H2.
 DR InterPro; IPR007086; Znf_C2H2_sub.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; zf-C2H2; 12.
 DR PRINTS; PR00048; ZINCFINGER.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 12.
 DR PROSITE; PS50805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 35
 FT ZN_FING 176 198 KRAB.
 FT ZN_FING 204 226 C2H2-TYPE 1 (DEGENERATE).
 FT ZN_FING 232 254 C2H2-TYPE 2.
 FT ZN_FING 260 282 C2H2-TYPE 3.
 FT ZN_FING 288 310 C2H2-TYPE 4.
 FT ZN_FING 316 338 C2H2-TYPE 5.
 FT ZN_FING 344 366 C2H2-TYPE 6.
 FT ZN_FING 372 394 C2H2-TYPE 7.
 FT ZN_FING 400 422 C2H2-TYPE 8.
 FT ZN_FING 428 450 C2H2-TYPE 9.
 FT ZN_FING 458 480 C2H2-TYPE 10.


```

FT   ZN_FING      456      478      C2H2-TYPE 11.
FT   ZN_FING      484      506      C2H2-TYPE 12.
FT   ZN_FING      512      534      C2H2-TYPE 13.
FT   ZN_FING      198      198      H -> R (IN REF. 2).
SQ   SEQUENCE      576 AA; 67216 MW; 532774BF69EC9E2A CRC64;

Query Match      17.3%; Score 385; DB 1; Length 576;
Best Local Similarity 29.4%; Pred. No. 7.5e-16;
Matches 109; Conservative 48; Mismatches 158; Indels 56; Gaps 15;

QY 8 KSISLSSSSSS-----RPKKYICTEGGCDKAYNRPSLLEQHLRTSHNDPRYKCTVDDCD 63
DB 211 KSPCMLHLSQHRIHIRENSYOC--EECGKAKWMTSLTRHKRIHIGKEKPKC--EECG 266
QY 64 KAFPRKSHLETHIVSHSEKPKPHSCVCGKQVNSRQHLKHEITHT--KSPKCTFENCQEA 121
DB 267 KAFKQSTLTTHKRIHIGKEKPKRCBEGKAFNRSSNLTTHKRIHIGKEKPKC--EEGKA 324
QY 122 FYKHQSI--RHILSVHEKTLTKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNF 180
DB 325 FNSOSTLSTHKFTIHAGEKPKYCECDKAFNRFTYLTKKRI--HIGKEKSYCEE--CGKG 381
QY 181 QTSVVLQFH--IKQSHPKLKCPCQKGCYVKGKGLSSHMLSHDSTWIKITWTCYDVGKF 238
DB 382 NMSSTLTTHKRIHIGKEKPKYCEVCGKAFNRSSNLTTHKRIHIGKEKPKC--GKA 436
QY 239 AKKNELVEHYNIPIHGNIPDDLKETEYVKLENIIDGSKLNLHLETEKLYEEDDED 298
DB 437 FNRSPQTLAKRIHTGK-----KPYKCEBGKAFSSSILTT----- 473
QY 299 EEDSLDEKRDVANDSQAQRSTKSFVSLSEGSVSKLISNSGKLTNCPKNCDFRMSF 358
DB 474 -----HKRIHIGKEKPKYCEBGKAFNRSSNLTTHKRIHIGKEKSYKC--EEGKAFNQ 523
QY 359 EYDLRRLKWH 369
DB 524 SSTLTTRKRIH 534

RESULT 11
2228 HUMAN STANDARD; PRT; 913 AA.
ID 090UT3; O9HCA7;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 228.
GN ZNF228.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Stubbbs L.,
RT "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.",
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kodoyanni V., Ge Y., Krummel G.K., Kvistad E., Grable L.,
RA Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.,
RT "Sequence analysis of a 1kb region in 19q13.2 containing a zinc finger
RT gene cluster.",
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUPEBEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 1 KRAA domain.
CC -!- SIMILARITY: Contains 17 C2H2-type zinc fingers.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF198358; AAF12816.1; -
DR EMBL; AC084239; AAG23968.1; -
DR HSSP; P08047; ISP2.
DR GeneW; HGNC:13021; ZNF228.
DR InterPro; IPR001909; KRAA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAA; 1.
DR Pfam; PF00096; ZF-C2H2; 13.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 12.
DR SMART; SM00349; KRAA; 1.
DR SMART; SM00355; ZNF_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 17.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein.
FT DOMAIN 8 79 KRAA.
FT ZN_FING 258 280 C2H2-TYPE 1 (DEGENERATE).
FT ZN_FING 443 465 C2H2-TYPE 2 (DEGENERATE).
FT ZN_FING 471 493 C2H2-TYPE 3 (DEGENERATE).
FT ZN_FING 497 519 C2H2-TYPE 4 (DEGENERATE).
FT ZN_FING 525 547 C2H2-TYPE 5.
FT ZN_FING 553 575 C2H2-TYPE 6.
FT ZN_FING 581 603 C2H2-TYPE 7.
FT ZN_FING 609 631 C2H2-TYPE 8.
FT ZN_FING 637 659 C2H2-TYPE 9.
FT ZN_FING 665 687 C2H2-TYPE 10.
FT ZN_FING 693 715 C2H2-TYPE 11.
FT ZN_FING 721 743 C2H2-TYPE 12.
FT ZN_FING 749 771 C2H2-TYPE 13.
FT ZN_FING 777 799 C2H2-TYPE 14.
FT ZN_FING 805 827 C2H2-TYPE 15.
FT ZN_FING 833 855 C2H2-TYPE 16.
FT ZN_FING 861 883 C2H2-TYPE 17.
FT ZN_FING 883 883 VSK -> KFO (IN REF. 2).
FT CONFLICT 3 113 T -> A (IN REF. 2).
FT CONFLICT 143 143 A -> V (IN REF. 2).
FT CONFLICT 222 222 E -> K (IN REF. 2).
FT CONFLICT 261 261 S -> T (IN REF. 2).
FT CONFLICT 392 392 Q -> N (IN REF. 2).
FT CONFLICT 446 446 Q -> E (IN REF. 2).
FT CONFLICT 485 485 H -> Y (IN REF. 2).
FT CONFLICT 813 813 A -> G (IN REF. 2).
SQ SEQUENCE 913 AA; 105702 MW; 8D19F5D27CD9F6B CRC64;

Query Match      17.2%; Score 383; DB 1; Length 913;
Best Local Similarity 30.7%; Pred. No. 1.6e-15;
Matches 109; Conservative 55; Mismatches 133; Indels 58; Gaps 19;

QY 23 KKYICTYEGCDKAYNRPSLLEQHLRTSHNDPRYKCTVDDCDKAFPRKSHLETHIVSHSEK 82
DB 551 KPYKC--EECDKGRSSSYLAQHRVHTGKEPKC--EECGKGFNRNSYLGQHQVRHTGE 606
QY 83 KPFHSCVCGKQVNSRQHLKHEITHT--KSPKCTFENCQEAFFYKHQSLR--HHILSVHEKT 139
DB 607 KPYKCEBGKAFNRSSNLTQHQRVHTGKEPKC--EECGKGFNRNSYLGQHQVRHTGKEP 664
QY 140 LTCQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNFQTSVVLQFH--IKQSHPKL 197
DB 665 YKCECGKGFSGASTLAKHQVRHTGKEP--YCQDE--CGKSFQSRSYLQSHQSVHSGERP 721
QY 198 KQPKQKGCYVKGKGLSSHMLSHDSTWIKITWTCYDVGKPAKKNELVEHYNIPIHGNIP 257

```



```

Db      722 ICEVCKGSGFSQRAVLGHQGRVH--TRVRYKCKEMCGKG--FSQSGRIEAHRRV-HTGGKP
Qy      258 DDLIKETEYVKLLENLLDQGSKLNNLHELETEKLVKDEDEDEEDSLDEKRSVPSDSMSA 317
Db      777 -----YKCE-VCTKGPS-----ESSRLQAHQ-----RVHVEGRPYKC 807
Qy      318 QRSIKSFT--ASLESGSKSVKLSNSGKK-INCPKNNCDMPREYDLRRHLKMH 369
Db      808 EGGCKAFSGSYSLQAHRRV-----HTGEKPYKC--EVCGKGSQSRNLTQAHQGRVH 855

RESULT 12
Z430 HUMAN
ID      Z430 HUMAN      STANDARD;      PRT;      570 AA.
AC      Q9H8G1; Q86V70;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Zinc finger protein 430.
GN      ZNF430.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini, Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Bi A., Yu L.;
RT      "Homo sapiens mRNA similar to zinc finger protein 85.";
RL      Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
RC      [2]
RE      SEQUENCE OF 1-364 FROM N.A.
RS      TTSUE-placenta;
RA      Itogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA      Nishikawa T., Nagai K., Sato H., Sugano S., Shintori A., Sudo H.,
RA      Wagaatsuma M., Hosoi T., Kaku Y., Kodaïra H., Kondo H., Sugawara M.,
RA      Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA      Matenabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA      Yamamoto U., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA      Nimomiya K., Iwayanagi T.;
RT      "NEDO human cDNA sequencing project.";
RL      Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: May be involved in transcriptional regulation.
CC      -!- SUBCELLULAR LOCATION: Nucleus (Potential).
CC      -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC      FINGER PROTEINS.
CC      -!- SIMILARITY: Contains 1 KRAB domain.
CC      -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AY269787; AAP30885.1; -
DR      EMBL; AK023721; BAB14656.1; -
DR      Genew: HGNC:20808; ZNF430.
DR      InterPro; IPR001909; KRAB.
DR      InterPro; IPR007087; Znf_C2H2.
DR      Pfam; PF01352; KRAB; 1.
DR      Pfam; PF00096; zf_C2H2; 3.
DR      ProDom; PD000003; Znf_C2H2; 3.
DR      SMART; SMO0349; KRAB_1.
DR      SMART; SMO0355; Znf_C2H2; 3.
DR      PROSITE; PS50805; KRAB; 1.
DR      PROSITE; PS50028; ZINC_FINGER_C2H2_1; 11.
DR      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
DR      Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
DR      Nuclear protein; Repeat.
DR      DOMAIN
DR      ZN_FING      35      107      KRAB
DR      ZN_FING      205      227      C2H2-TYPE 1.

```

FT	ZN_FING	233	255	CZH2-TYPE 2.
FT	ZN_FING	261	283	CZH2-TYPE 3.
FT	ZN_FING	289	311	CZH2-TYPE 4.
FT	ZN_FING	317	339	CZH2-TYPE 5.
FT	ZN_FING	345	367	CZH2-TYPE 6.
FT	ZN_FING	373	395	CZH2-TYPE 7.
FT	ZN_FING	401	423	CZH2-TYPE 8.
FT	ZN_FING	429	451	CZH2-TYPE 9.
FT	ZN_FING	457	479	CZH2-TYPE 10.
FT	ZN_FING	485	507	CZH2-TYPE 11.
FT	ZN_FING	513	535	CZH2-TYPE 12 (DEGENERATE).
FT	CONFLICT	293	364	EEGKTPNRRSSHTLTHKRIHTGKPYRCECGKAFNRSSHT
FT				THKLIHTGKPYRCECGKAFNRSSHTLTHKRIHTGKPYRCECGKAFNRSSHT
FT				TEPHTLTHKEPILEPRNPDPVAGVAELTGPHTLTHIRLEI
FT				LERNPDPVAGVAELTGPHTLTHIR (IN REF. 2).
SQ	SEQUENCE	570 AA;	66346 MW;	0561EB871DC92B61F CRC64;
Query Match		16.9%;	Score 376;	DB 1; Length 570;
Best Local Similarity		32.5%;	Pred. No. 2.5e-15;	
Matches 94;	Conservative	38;	Mismatches 109;	Indels 48; Gaps 13
QY	23	KKYLCTYECCDQAVNRPRLLEOHLRTHSNDRYKCTGVDDCDCAFPRKSLTHRIYSHSK	82	
DB	287	KKYRC--EECGTTPNRSSHTLTHKRIHTGKPYRC--EECGAFNRSSHTLTHKRIHTGE	342	
QY	83	KPHCSVCGKGVNSBOHLKRHEIHT--KSPKCTEENCQEAFFYKQSL-RHILISVHEK	138	
DB	343	KPKYCECGKAFNQSLTTHKLIHAGKPKYC--EECGKAFYRSYLTAKLIHTGKPF	400	
QY	139	-----TLT-----CKCNKVTTPRSKLAQHKLKHGGSPAYQC	171	
DB	401	YKCEECGKGFNWSLTTHKRIHTGKPYRCECGKAFNRSSHTLTHKRIHTGKPF-YKC	459	
QY	172	DHGEFGKQFTWSVYQFH--IKQSHPKLCKPKCGKCGCGKGLSHMLSHDSTMTIKTWT	229	
DB	460	EE--CGKAFNRSPLTHKRIHTGKPYRCECGKAFNQSLTTHKLIHTGKPF	517	
QY	230	CDYCDVGFAPKAKNLEVEHYNIFFHDGNTIPDDLKETEY-KKLEINLDQGS	277	
DB	518	CDKA---FSQSSLTTHK-KVLIHGEKRYNEEYKGAHQNSNLTQGSN	561	
RESULT 13				
ID	2184_HUMAN			
AC	Q99676; O60792;	STANDARD;	PRT;	751 AA.
DT	15-JUL-1998 (Rel. 36, Created)			
DT	16-OCT-2001 (Rel. 40, last sequence update)			
DT	10-OCT-2003 (Rel. 42, last annotation update)			
DE	Zinc finger protein 184.			
GN	ZNF184.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RA	SEQUENCE FROM N.A.			
RA	Phillips S.;			
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RN	SEQUENCE OF 26-751 FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=97230463; PubMed=9073517;			
RA	Goldwurm S., Menzies M.L., Banger J.L., Powell B.L.W.,			
RA	Jazwinska E.C.;			
RT	"Identification of a novel Krueppel-related zinc finger gene (ZNF184)			
RT	mapping to 6p21.3."/.			
CC	Genomics 40:486-489(1997).			
CC	-1- FUNCTION: May be involved in transcriptional regulation.			
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-1- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS.			
CC	-1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF CZH2-TYPE ZINC-FINGER PROTEINS.			

```

CC - SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL021918; CAAL7278.1; -.
DR EMBL: U66561; AAC51180.1; -.
DR HSSP: P08046; 1A1H; ZNF184.
DR Genew: HGNC:12975; ZNF184.
DR MIM: 602277; -.
DR GO: GO:0005634; C:nucleus; NAS.
DR GO: GO:0003677; F:DNA binding; NAS.
DR GO: GO:0008270; F:zinc ion binding; NAS.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR007087; Znf_C2H2.
DR InterPro: IPR007086; Znf_C2H2_sub.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; zf-C2H2; 19.
DR PRINTS: PRO0048; ZINC_FINGER.
DR ProDom: PD000003; Znf_C2H2; 18.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 19.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 19.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 19.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
KW DOMAIN 28 99
FT ZN_FING 222 244 KRAB-TYPE.
FT ZN_FING 250 272 C2H2-TYPE.
FT ZN_FING 278 300 C2H2-TYPE.
FT ZN_FING 306 328 C2H2-TYPE.
FT ZN_FING 334 356 C2H2-TYPE.
FT ZN_FING 362 384 C2H2-TYPE.
FT ZN_FING 390 412 C2H2-TYPE.
FT ZN_FING 418 440 C2H2-TYPE.
FT ZN_FING 446 468 C2H2-TYPE.
FT ZN_FING 474 496 C2H2-TYPE.
FT ZN_FING 502 524 C2H2-TYPE.
FT ZN_FING 530 552 C2H2-TYPE.
FT ZN_FING 558 580 C2H2-TYPE.
FT ZN_FING 586 608 C2H2-TYPE.
FT ZN_FING 614 636 C2H2-TYPE.
FT ZN_FING 642 664 C2H2-TYPE.
FT ZN_FING 670 692 C2H2-TYPE.
FT ZN_FING 698 720 C2H2-TYPE.
FT ZN_FING 726 748 C2H2-TYPE.
FT ZN_FING 754 776 C2H2-TYPE.
FT ZN_FING 782 804 C2H2-TYPE.
FT ZN_FING 810 832 C2H2-TYPE.
FT ZN_FING 838 860 C2H2-TYPE.
FT ZN_FING 868 890 C2H2-TYPE.
FT ZN_FING 898 920 C2H2-TYPE.
FT ZN_FING 926 948 C2H2-TYPE.
FT ZN_FING 954 976 C2H2-TYPE.
FT ZN_FING 982 1004 C2H2-TYPE.
FT ZN_FING 1010 1032 C2H2-TYPE.
FT ZN_FING 1038 1060 C2H2-TYPE.
FT ZN_FING 1068 1090 C2H2-TYPE.
FT ZN_FING 1098 1120 C2H2-TYPE.
FT ZN_FING 1128 1150 C2H2-TYPE.
FT ZN_FING 1158 1180 C2H2-TYPE.
FT ZN_FING 1188 1210 C2H2-TYPE.
FT ZN_FING 1218 1240 C2H2-TYPE.
FT ZN_FING 1248 1270 C2H2-TYPE.
FT ZN_FING 1278 1300 C2H2-TYPE.
FT ZN_FING 1308 1330 C2H2-TYPE.
FT ZN_FING 1338 1360 C2H2-TYPE.
FT ZN_FING 1368 1390 C2H2-TYPE.
FT ZN_FING 1398 1420 C2H2-TYPE.
FT ZN_FING 1428 1450 C2H2-TYPE.
FT ZN_FING 1458 1480 C2H2-TYPE.
FT ZN_FING 1488 1510 C2H2-TYPE.
FT ZN_FING 1518 1540 C2H2-TYPE.
FT ZN_FING 1548 1570 C2H2-TYPE.
FT ZN_FING 1578 1600 C2H2-TYPE.
FT ZN_FING 1608 1630 C2H2-TYPE.
FT ZN_FING 1638 1660 C2H2-TYPE.
FT ZN_FING 1668 1690 C2H2-TYPE.
FT ZN_FING 1698 1720 C2H2-TYPE.
FT ZN_FING 1728 1750 C2H2-TYPE.
FT ZN_FING 1758 1780 C2H2-TYPE.
FT ZN_FING 1788 1810 C2H2-TYPE.
FT ZN_FING 1818 1840 C2H2-TYPE.
FT ZN_FING 1848 1870 C2H2-TYPE.
FT ZN_FING 1878 1900 C2H2-TYPE.
FT ZN_FING 1908 1930 C2H2-TYPE.
FT ZN_FING 1938 1960 C2H2-TYPE.
FT ZN_FING 1968 1990 C2H2-TYPE.
FT ZN_FING 1998 2020 C2H2-TYPE.
FT ZN_FING 2028 2050 C2H2-TYPE.
FT ZN_FING 2058 2080 C2H2-TYPE.
FT ZN_FING 2088 2110 C2H2-TYPE.
FT ZN_FING 2118 2140 C2H2-TYPE.
FT ZN_FING 2148 2170 C2H2-TYPE.
FT ZN_FING 2178 2200 C2H2-TYPE.
FT ZN_FING 2208 2230 C2H2-TYPE.
FT ZN_FING 2238 2260 C2H2-TYPE.
FT ZN_FING 2268 2290 C2H2-TYPE.
FT ZN_FING 2298 2320 C2H2-TYPE.
FT ZN_FING 2328 2350 C2H2-TYPE.
FT ZN_FING 2358 2380 C2H2-TYPE.
FT ZN_FING 2388 2410 C2H2-TYPE.
FT ZN_FING 2418 2440 C2H2-TYPE.
FT ZN_FING 2448 2470 C2H2-TYPE.
FT ZN_FING 2478 2500 C2H2-TYPE.
FT ZN_FING 2508 2530 C2H2-TYPE.
FT ZN_FING 2538 2560 C2H2-TYPE.
FT ZN_FING 2568 2590 C2H2-TYPE.
FT ZN_FING 2598 2620 C2H2-TYPE.
FT ZN_FING 2628 2650 C2H2-TYPE.
FT ZN_FING 2658 2680 C2H2-TYPE.
FT ZN_FING 2688 2710 C2H2-TYPE.
FT ZN_FING 2718 2740 C2H2-TYPE.
FT ZN_FING 2748 2770 C2H2-TYPE.
FT ZN_FING 2778 2800 C2H2-TYPE.
FT ZN_FING 2808 2830 C2H2-TYPE.
FT ZN_FING 2838 2860 C2H2-TYPE.
FT ZN_FING 2868 2890 C2H2-TYPE.
FT ZN_FING 2898 2920 C2H2-TYPE.
FT ZN_FING 2928 2950 C2H2-TYPE.
FT ZN_FING 2958 2980 C2H2-TYPE.
FT ZN_FING 2988 3010 C2H2-TYPE.
FT ZN_FING 3018 3040 C2H2-TYPE.
FT ZN_FING 3048 3070 C2H2-TYPE.
FT ZN_FING 3078 3100 C2H2-TYPE.
FT ZN_FING 3108 3130 C2H2-TYPE.
FT ZN_FING 3138 3160 C2H2-TYPE.
FT ZN_FING 3168 3190 C2H2-TYPE.
FT ZN_FING 3198 3220 C2H2-TYPE.
FT ZN_FING 3228 3250 C2H2-TYPE.
FT ZN_FING 3258 3280 C2H2-TYPE.
FT ZN_FING 3288 3310 C2H2-TYPE.
FT ZN_FING 3318 3340 C2H2-TYPE.
FT ZN_FING 3348 3370 C2H2-TYPE.
FT ZN_FING 3378 3400 C2H2-TYPE.
FT ZN_FING 3408 3430 C2H2-TYPE.
FT ZN_FING 3438 3460 C2H2-TYPE.
FT ZN_FING 3468 3490 C2H2-TYPE.
FT ZN_FING 3498 3520 C2H2-TYPE.
FT ZN_FING 3528 3550 C2H2-TYPE.
FT ZN_FING 3558 3580 C2H2-TYPE.
FT ZN_FING 3588 3610 C2H2-TYPE.
FT ZN_FING 3618 3640 C2H2-TYPE.
FT ZN_FING 3648 3670 C2H2-TYPE.
FT ZN_FING 3678 3700 C2H2-TYPE.
FT ZN_FING 3708 3730 C2H2-TYPE.
FT ZN_FING 3738 3760 C2H2-TYPE.
FT ZN_FING 3768 3790 C2H2-TYPE.
FT ZN_FING 3798 3820 C2H2-TYPE.
FT ZN_FING 3828 3850 C2H2-TYPE.
FT ZN_FING 3858 3880 C2H2-TYPE.
FT ZN_FING 3888 3910 C2H2-TYPE.
FT ZN_FING 3918 3940 C2H2-TYPE.
FT ZN_FING 3948 3970 C2H2-TYPE.
FT ZN_FING 3978 4000 C2H2-TYPE.
FT ZN_FING 4008 4030 C2H2-TYPE.
FT ZN_FING 4038 4060 C2H2-TYPE.
FT ZN_FING 4068 4090 C2H2-TYPE.
FT ZN_FING 4098 4120 C2H2-TYPE.
FT ZN_FING 4128 4150 C2H2-TYPE.
FT ZN_FING 4158 4180 C2H2-TYPE.
FT ZN_FING 4188 4210 C2H2-TYPE.
FT ZN_FING 4218 4240 C2H2-TYPE.
FT ZN_FING 4248 4270 C2H2-TYPE.
FT ZN_FING 4278 4300 C2H2-TYPE.
FT ZN_FING 4308 4330 C2H2-TYPE.
FT ZN_FING 4338 4360 C2H2-TYPE.
FT ZN_FING 4368 4390 C2H2-TYPE.
FT ZN_FING 4398 4420 C2H2-TYPE.
FT ZN_FING 4428 4450 C2H2-TYPE.
FT ZN_FING 4458 4480 C2H2-TYPE.
FT ZN_FING 4488 4510 C2H2-TYPE.
FT ZN_FING 4518 4540 C2H2-TYPE.
FT ZN_FING 4548 4570 C2H2-TYPE.
FT ZN_FING 4578 4600 C2H2-TYPE.
FT ZN_FING 4608 4630 C2H2-TYPE.
FT ZN_FING 4638 4660 C2H2-TYPE.
FT ZN_FING 4668 4690 C2H2-TYPE.
FT ZN_FING 4698 4720 C2H2-TYPE.
FT ZN_FING 4728 4750 C2H2-TYPE.
FT ZN_FING 4758 4780 C2H2-TYPE.
FT ZN_FING 4788 4810 C2H2-TYPE.
FT ZN_FING 4818 4840 C2H2-TYPE.
FT ZN_FING 4848 4870 C2H2-TYPE.
FT ZN_FING 4878 4900 C2H2-TYPE.
FT ZN_FING 4908 4930 C2H2-TYPE.
FT ZN_FING 4938 4960 C2H2-TYPE.
FT ZN_FING 4968 4990 C2H2-TYPE.
FT ZN_FING 4998 5020 C2H2-TYPE.
FT ZN_FING 5028 5050 C2H2-TYPE.
FT ZN_FING 5058 5080 C2H2-TYPE.
FT ZN_FING 5088 5110 C2H2-TYPE.
FT ZN_FING 5118 5140 C2H2-TYPE.
FT ZN_FING 5148 5170 C2H2-TYPE.
FT ZN_FING 5178 5200 C2H2-TYPE.
FT ZN_FING 5208 5230 C2H2-TYPE.
FT ZN_FING 5238 5260 C2H2-TYPE.
FT ZN_FING 5268 5290 C2H2-TYPE.
FT ZN_FING 5298 5320 C2H2-TYPE.
FT ZN_FING 5328 5350 C2H2-TYPE.
FT ZN_FING 5358 5380 C2H2-TYPE.
FT ZN_FING 5388 5410 C2H2-TYPE.
FT ZN_FING 5418 5440 C2H2-TYPE.
FT ZN_FING 5448 5470 C2H2-TYPE.
FT ZN_FING 5478 5500 C2H2-TYPE.
FT ZN_FING 5508 5530 C2H2-TYPE.
FT ZN_FING 5538 5560 C2H2-TYPE.
FT ZN_FING 5568 5590 C2H2-TYPE.
FT ZN_FING 5598 5620 C2H2-TYPE.
FT ZN_FING 5628 5650 C2H2-TYPE.
FT ZN_FING 5658 5680 C2H2-TYPE.
FT ZN_FING 5688 5710 C2H2-TYPE.
FT ZN_FING 5718 5740 C2H2-TYPE.
FT ZN_FING 5748 5770 C2H2-TYPE.
FT ZN_FING 5778 5800 C2H2-TYPE.
FT ZN_FING 5808 5830 C2H2-TYPE.
FT ZN_FING 5838 5860 C2H2-TYPE.
FT ZN_FING 5868 5890 C2H2-TYPE.
FT ZN_FING 5898 5920 C2H2-TYPE.
FT ZN_FING 5928 5950 C2H2-TYPE.
FT ZN_FING 5958 5980 C2H2-TYPE.
FT ZN_FING 5988 6010 C2H2-TYPE.
FT ZN_FING 6018 6040 C2H2-TYPE.
FT ZN_FING 6048 6070 C2H2-TYPE.
FT ZN_FING 6078 6100 C2H2-TYPE.
FT ZN_FING 6108 6130 C2H2-TYPE.
FT ZN_FING 6138 6160 C2H2-TYPE.
FT ZN_FING 6168 6190 C2H2-TYPE.
FT ZN_FING 6198 6220 C2H2-TYPE.
FT ZN_FING 6228 6250 C2H2-TYPE.
FT ZN_FING 6258 6280 C2H2-TYPE.
FT ZN_FING 6288 6310 C2H2-TYPE.
FT ZN_FING 6318 6340 C2H2-TYPE.
FT ZN_FING 6348 6370 C2H2-TYPE.
FT ZN_FING 6378 6400 C2H2-TYPE.
FT ZN_FING 6408 6430 C2H2-TYPE.
FT ZN_FING 6438 6460 C2H2-TYPE.
FT ZN_FING 6468 6490 C2H2-TYPE.
FT ZN_FING 6498 6520 C2H2-TYPE.
FT ZN_FING 6528 6550 C2H2-TYPE.
FT ZN_FING 6558 6580 C2H2-TYPE.
FT ZN_FING 6588 6610 C2H2-TYPE.
FT ZN_FING 6618 6640 C2H2-TYPE.
FT ZN_FING 6648 6670 C2H2-TYPE.
FT ZN_FING 6678 6700 C2H2-TYPE.
FT ZN_FING 6708 6730 C2H2-TYPE.
FT ZN_FING 6738 6760 C2H2-TYPE.
FT ZN_FING 6768 6790 C2H2-TYPE.
FT ZN_FING 6798 6820 C2H2-TYPE.
FT ZN_FING 6828 6850 C2H2-TYPE.
FT ZN_FING 6858 6880 C2H2-TYPE.
FT ZN_FING 6888 6910 C2H2-TYPE.
FT ZN_FING 6918 6940 C2H2-TYPE.
FT ZN_FING 6948 6970 C2H2-TYPE.
FT ZN_FING 6978 7000 C2H2-TYPE.
FT ZN_FING 7008 7030 C2H2-TYPE.
FT ZN_FING 7038 7060 C2H2-TYPE.
FT ZN_FING 7068 7090 C2H2-TYPE.
FT ZN_FING 7098 7120 C2H2-TYPE.
FT ZN_FING 7128 7150 C2H2-TYPE.
FT ZN_FING 7158 7180 C2H2-TYPE.
FT ZN_FING 7188 7210 C2H2-TYPE.
FT ZN_FING 7218 7240 C2H2-TYPE.
FT ZN_FING 7248 7270 C2H2-TYPE.
FT ZN_FING 7278 7300 C2H2-TYPE.
FT ZN_FING 7308 7330 C2H2-TYPE.
FT ZN_FING 7338 7360 C2H2-TYPE.
FT ZN_FING 7368 7390 C2H2-TYPE.
FT ZN_FING 7398 7420 C2H2-TYPE.
FT ZN_FING 7428 7450 C2H2-TYPE.
FT ZN_FING 7458 7480 C2H2-TYPE.
FT ZN_FING 7488 7510 C2H2-TYPE.
FT ZN_FING 7518 7540 C2H2-TYPE.
FT ZN_FING 7548 7570 C2H2-TYPE.
FT ZN_FING 7578 7600 C2H2-TYPE.
FT ZN_FING 7608 7630 C2H2-TYPE.
FT ZN_FING 7638 7660 C2H2-TYPE.
FT ZN_FING 7668 7690 C2H2-TYPE.
FT ZN_FING 7698 7720 C2H2-TYPE.
FT ZN_FING 7728 7750 C2H2-TYPE.
FT ZN_FING 7758 7780 C2H2-TYPE.
FT ZN_FING 7788 7810 C2H2-TYPE.
FT ZN_FING 7818 7840 C2H2-TYPE.
FT ZN_FING 7848 7870 C2H2-TYPE.
FT ZN_FING 7878 7900 C2H2-TYPE.
FT ZN_FING 7908 7930 C2H2-TYPE.
FT ZN_FING 7938 7960 C2H2-TYPE.
FT ZN_FING 7968 7990 C2H2-TYPE.
FT ZN_FING 7998 8020 C2H2-TYPE.
FT ZN_FING 8028 8050 C2H2-TYPE.
FT ZN_FING 8058 8080 C2H2-TYPE.
FT ZN_FING 8088 8110 C2H2-TYPE.
FT ZN_FING 8118 8140 C2H2-TYPE.
FT ZN_FING 8148 8170 C2H2-TYPE.
FT ZN_FING 8178 8200 C2H2-TYPE.
FT ZN_FING 8208 8230 C2H2-TYPE.
FT ZN_FING 8238 8260 C2H2-TYPE.
FT ZN_FING 8268 8290 C2H2-TYPE.
FT ZN_FING 8298 8320 C2H2-TYPE.
FT ZN_FING 8328 8350 C2H2-TYPE.
FT ZN_FING 8358 8380 C2H2-TYPE.
FT ZN_FING 8388 8410 C2H2-TYPE.
FT ZN_FING 8418 8440 C2H2-TYPE.
FT ZN_FING 8448 8470 C2H2-TYPE.
FT ZN_FING 8478 8500 C2H2-TYPE.
FT ZN_FING 8508 8530 C2H2-TYPE.
FT ZN_FING 8538 8560 C2H2-TYPE.
FT ZN_FING 8568 8590 C2H2-TYPE.
FT ZN_FING 8598 8620 C2H2-TYPE.
FT ZN_FING 8628 8650 C2H2-TYPE.
FT ZN_FING 8658 8680 C2H2-TYPE.
FT ZN_FING 8688 8710 C2H2-TYPE.
FT ZN_FING 8718 8740 C2H2-TYPE.
FT ZN_FING 8748 8770 C2H2-TYPE.
FT ZN_FING 8778 8800 C2H2-TYPE.
FT ZN_FING 8808 8830 C2H2-TYPE.
FT ZN_FING 8838 8860 C2H2-TYPE.
FT ZN_FING 8868 8890 C2H2-TYPE.
FT ZN_FING 8898 8920 C2H2-TYPE.
FT ZN_FING 8928 8950 C2H2-TYPE.
FT ZN_FING 8958 8980 C2H2-TYPE.
FT ZN_FING 8988 9010 C2H2-TYPE.
FT ZN_FING 9018 9040 C2H2-TYPE.
FT ZN_FING 9048 9070 C2H2-TYPE.
FT ZN_FING 9078 9100 C2H2-TYPE.
FT ZN_FING 9108 9130 C2H2-TYPE.
FT ZN_FING 9138 9160 C2H2-TYPE.
FT ZN_FING 9168 9190 C2H2-TYPE.
FT ZN_FING 9198 9220 C2H2-TYPE.
FT ZN_FING 9228 9250 C2H2-TYPE.
FT ZN_FING 9258 9280 C2H2-TYPE.
FT ZN_FING 9288 9310 C2H2-TYPE.
FT ZN_FING 9318 9340 C2H2-TYPE.
FT ZN_FING 9348 9370 C2H2-TYPE.
FT ZN_FING 9378 9400 C2H2-TYPE.
FT ZN_FING 9408 9430 C2H2-TYPE.
FT ZN_FING 9438 9460 C2H2-TYPE.
FT ZN_FING 9468 9490 C2H2-TYPE.
FT ZN_FING 9498 9520 C2H2-TYPE.
FT ZN_FING 9528 9550 C2H2-TYPE.
FT ZN_FING 9558 9580 C2H2-TYPE.
FT ZN_FING 9588 9610 C2H2-TYPE.
FT ZN_FING 9618 9640 C2H2-TYPE.
FT ZN_FING 9648 9670 C2H2-TYPE.
FT ZN_FING 9678 9700 C2H2-TYPE.
FT ZN_FING 9708 9730 C2H2-TYPE.
FT ZN_FING 9738 9760 C2H2-TYPE.
FT ZN_FING 9768 9790 C2H2-TYPE.
FT ZN_FING 9798 9820 C2H2-TYPE.
FT ZN_FING 9828 9850 C2H2-TYPE.
FT ZN_FING 9858 9880 C2H2-TYPE.
FT ZN_FING 9888 9910 C2H2-TYPE.
FT ZN_FING 9918 9940 C2H2-TYPE.
FT ZN_FING 9948 9970 C2H2-TYPE.
FT ZN_FING 9978 10000 C2H2-TYPE.
FT ZN_FING 10008 10030 C2H2-TYPE.
FT ZN_FING 10038 10060 C2H2-TYPE.
FT ZN_FING 10068 10090 C2H2-TYPE.
FT ZN_FING 10098 10120 C2H2-TYPE.
FT ZN_FING 10128 10150 C2H2-TYPE.
FT ZN_FING 10158 10180 C2H2-TYPE.
FT ZN_FING 10188 10210 C2H2-TYPE.
FT ZN_FING 10218 10240 C2H2-TYPE.
FT ZN_FING 10248 10270 C2H2-TYPE.
FT ZN_FING 10278 10300 C2H2-TYPE.
FT ZN_FING 10308 10330 C2H2-TYPE.
FT ZN_FING 10338 10360 C2H2-TYPE.
FT ZN_FING 10368 10390 C2H2-TYPE.
FT ZN_FING 10398 10420 C2H2-TYPE.
FT ZN_FING 10428 10450 C2H2-TYPE.
FT ZN_FING 10458 10480 C2H2-TYPE.
FT ZN_FING 10488 10510 C2H2-TYPE.
FT ZN_FING 10518 10540 C2H2-TYPE.
FT ZN_FING 10548 10570 C2H2-TYPE.
FT ZN_FING 10578 10600 C2H2-TYPE.
FT ZN_FING 10608 10630 C2H2-TYPE.
FT ZN_FING 10638 10660 C2H2-TYPE.
FT ZN_FING 10668 10690 C2H2-TYPE.
FT ZN_FING 10698 10720 C2H2-TYPE.
FT ZN_FING 10728 10750 C2H2-TYPE.
FT ZN_FING 10758 10780 C2H2-TYPE.
FT ZN_FING 10788 10810 C2H2-TYPE.
FT ZN_FING 10818 10840 C2H2-TYPE.
FT ZN_FING 10848 10870 C2H2-TYPE.
FT ZN_FING 10878 10900 C2H2-TYPE.
FT ZN_FING 10908 10930 C2H2-TYPE.
FT ZN_FING 10938 10960 C2H2-TYPE.
FT ZN_FING 10968 10990 C2H2-TYPE.
FT ZN_FING 10998 11020 C2H2-TYPE.
FT ZN_FING 11028 11050 C2H2-TYPE.
FT ZN_FING 11058 11080 C2H2-TYPE.
FT ZN_FING 11088 11110 C2H2-TYPE.
FT ZN_FING 11118 11140 C2H2-TYPE.
FT ZN_FING 11148 11170 C2H2-TYPE.
FT ZN_FING 11178 11200 C2H2-TYPE.
FT ZN_FING 11208 11230 C2H2-TYPE.
FT ZN_FING 11238 11260 C2H2-TYPE.
FT ZN_FING 11268 11290 C2H2-TYPE.
FT ZN_FING 11298 11320 C2H2-TYPE.
FT ZN_FING 11328 11350 C2H2-TYPE.
FT ZN_FING 11358 11380 C2H2-TYPE.
FT ZN_FING 11388 11410 C2H2-TYPE.
FT ZN_FING 11418 11440 C2H2-TYPE.
FT ZN_FING 11448 11470 C2H2-TYPE.
FT ZN_FING 11478 11500 C2H2-TYPE.
FT ZN_FING 11508 11530 C2H2-TYPE.
FT ZN_FING 11538 11560 C2H2-TYPE.
FT ZN_FING 11568 11590 C2H2-TYPE.
FT ZN_FING 11598 11620 C2H2-TYPE.
FT ZN_FING 11628 11650 C2H2-TYPE.
FT ZN_FING 11658 11680 C2H2-TYPE.
FT ZN_FING 11688 11710 C2H2-TYPE.
FT ZN_FING 11718 11740 C2H2-TYPE.
FT ZN_FING 11748 11770 C2H2-TYPE.
FT ZN_FING 11778 11800 C2H2-TYPE.
FT ZN_FING 11808 11830 C2H2-TYPE.
FT ZN_FING 11838 11860 C2H2-TYPE.
FT ZN_FING 11868 11890 C2H2-TYPE.
FT ZN_FING 11898 11920 C2H2-TYPE.
FT ZN_FING 11928 11950 C2H2-TYPE.
FT ZN_FING 11958 11980 C2H2-TYPE.
FT ZN_FING 11988 12010 C2H2-TYPE.
FT ZN_FING 12018 12040 C2H2-TYPE.
FT ZN_FING 12048 12070 C2H2-TYPE.
FT ZN_FING 12078 12100 C2H2-TYPE.
FT ZN_FING 12108 12130 C2H2-TYPE.
FT ZN_FING 12138 12160 C2H2-TYPE.
FT ZN_FING 12168 12190 C2H2-TYPE.
FT ZN_FING 12198 12220 C2H2-TYPE.
FT ZN_FING 12228 12250 C2H2-TYPE.
FT ZN_FING 12258 12280 C2H2-TYPE.
FT ZN_FING 12288 12310 C2H2-TYPE.
FT ZN_FING 12318 12340 C2H2-TYPE.
FT ZN_FING 12348 12370 C2H2-TYPE.
FT ZN_FING 12378 12400 C2H2-TYPE.
FT ZN_FING 12408 12430 C2H2-TYPE.
FT ZN_FING 12438 12460 C2H2-TYPE.
FT ZN_FING 12468 12490 C2H2-TYPE.
FT ZN_FING 12498 12520 C2H2-TYPE.
FT ZN_FING 12528 12550 C2H2-TYPE.
FT ZN_FING 12558 12580 C2H2-TYPE.
FT ZN_FING 12588 12610 C2H2-TYPE.
FT ZN_FING 12618 12640 C2H2-TYPE.
FT ZN_FING 12648 12670 C2H2-TYPE.
FT ZN_FING 12678 12700 C2H2-TYPE.
FT ZN_FING 12708 12730 C2H2-TYPE.
FT ZN_FING 12738 12760 C2H2-TYPE.
FT ZN_FING 12768 12790 C2H2-TYPE.
FT ZN_FING 12798 12820 C2H2-TYPE.
FT ZN_FING 12828 12850 C2H2-TYPE.
FT ZN_FING 12858 12880 C2H2-TYPE.
FT ZN_FING 12888 12910 C2H2-TYPE.
FT ZN_FING 12918 12940 C2H2-TYPE.
FT ZN_FING 12948 12970 C2H2-TYPE.
FT ZN_FING 12978 13000 C2H2-TYPE.
FT ZN_FING 13008 13030 C2H2-TYPE.
FT ZN_FING 13038 13060 C2H2-TYPE.
FT ZN_FING 13068 13090 C2H2-TYPE.
FT ZN_FING 13098 13120 C2H2-TYPE.
FT ZN_FING 13128 13150 C2H2-TYPE.
FT ZN_FING 13158 13180 C2H2-TYPE.
FT ZN_FING 13188 13210 C2H2-TYPE.
FT ZN_FING 13218 13240 C2H2-TYPE.
FT ZN_FING 13248 13270 C2H2-TYPE.
FT ZN_FING 13278 13300 C2H2-TYPE.
FT ZN_FING 13308 13330 C2H2-TYPE.
FT ZN_FING 13338 13360 C2H2-TYPE.
FT ZN_FING 13368 13390 C2H2-TYPE.
FT ZN_FING 13398 13420 C2H2-TYPE.
FT ZN_FING 13428 13450 C2H2-TYPE.
FT ZN_FING 13458 13480 C2H2-TYPE.
FT ZN_FING 13488 13510 C2H2-TYPE.
FT ZN_FING 13518 13540 C2H2-TYPE.
FT ZN_FING 13548 13570 C2H2-TYPE.
FT ZN_FING 13578 13600 C2H2-TYPE.
FT ZN_FING 13608 13630 C2H2-TYPE.
FT ZN_FING 13638 13660 C2H2-TYPE.
FT ZN_FING 13668 13690 C2H2-TYPE.
FT ZN_FING 13698 13720 C2H2-TYPE.
FT ZN_FING 13728 13750 C2H2-TYPE.
FT ZN_FING 13758 13780 C2H2-TYPE.
FT ZN_FING 13788 13810 C2H2-TYPE.
FT ZN_FING 13818 13840 C2H2-TYPE.
FT ZN_FING 13848 13870 C2H2-TYPE.
FT ZN_FING 13878 13900 C2H2-TYPE.
FT ZN_FING 13908 13930 C2H2-TYPE.
FT ZN_FING 13938 13960 C2H2-TYPE.
FT ZN_FING 13968 13990 C2H2-TYPE.
FT ZN_FING 13998 14020 C2H2-TYPE.
FT ZN_FING 14028 14050 C2H2-TYPE.
FT ZN_FING 14058 14080 C2H2-TYPE.
FT ZN_FING 14088 14110 C2H2-TYPE.
FT ZN_FING 14118 14140 C2H2-TYPE.
FT ZN_FING 14148 14170 C2H2-TYPE.
FT ZN_FING 14178 14200 C2H2-TYPE.
FT ZN_FING 14208 14230 C2H2-TYPE.
FT ZN_FING 14238 14260 C2H2-TYPE.
FT ZN_FING 14268 14290 C2H2-TYPE.
FT ZN_FING 14298 14320 C2H2-TYPE.
FT ZN_FING 14328 14350 C2H2-TYPE.
FT ZN_FING 14358 14380 C2H2-TYPE.
FT ZN_FING 14388 14410 C2H2-TYPE.
FT ZN_FING 14418 14440 C2H2-TYPE.
FT ZN_FING 14448 14470 C2H2-TYPE.
FT ZN_FING 14478 14500 C2H2-TYPE.
FT ZN_FING 14508 14530 C2H2-TYPE.
FT ZN_FING 14538 14560 C2H2-TYPE.
FT ZN_FING 14568 14590 C2H2-TYPE.
FT ZN_FING 14598 14620 C2H2-TYPE.
FT ZN_FING 14628 14650 C2H2-TYPE.
FT ZN_FING 14658 14680 C2H2-TYPE.
FT ZN_FING 14688 14710 C2H2-TYPE.
FT ZN_FING 14718 14740 C2H2-TYPE.
FT ZN_FING 14748 14770 C2H2-TYPE.
FT ZN_FING 14778 14800 C2H2-TYPE.
FT ZN_FING 14808 14830 C2H2-TYPE.
FT ZN_FING 14838 14860 C2H2-TYPE.
FT ZN_FING 14868 14890 C2H2-TYPE.
FT ZN_FING 14898 14920 C2H2-TYPE.
FT ZN_FING 14928 1495
```

CC REGULATION PROCESSES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
CC AS NEURAL RETINA CONES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
CC EMBRYOGENESIS.
CC -1- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06021; CAA29425.1; -.
DR PDB; 1ZNF; 15-OCT-91.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf_C2H2; 36.
DR ProDom; PD000003; Znf_C2H2; 20.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 35.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 35.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 37.
KW Zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;
KW Phosphorylation.
FT DOMAIN 1 58 KRAB.
FT ZN_FING 108 130 C2H2-TYPE.
FT ZN_FING 136 158 C2H2-TYPE.
FT ZN_FING 164 186 C2H2-TYPE.
FT ZN_FING 192 214 C2H2-TYPE.
FT ZN_FING 220 242 C2H2-TYPE.
FT ZN_FING 248 270 C2H2-TYPE.
FT ZN_FING 276 298 C2H2-TYPE.
FT ZN_FING 326 348 C2H2-TYPE.
FT ZN_FING 354 376 C2H2-TYPE.
FT ZN_FING 382 404 C2H2-TYPE.
FT ZN_FING 410 432 C2H2-TYPE.
FT ZN_FING 438 460 C2H2-TYPE.
FT ZN_FING 466 488 C2H2-TYPE.
FT ZN_FING 503 525 C2H2-TYPE.
FT ZN_FING 531 553 C2H2-TYPE.
FT ZN_FING 559 581 C2H2-TYPE.
FT ZN_FING 587 609 C2H2-TYPE.
FT ZN_FING 615 637 C2H2-TYPE.
FT ZN_FING 643 665 C2H2-TYPE.
FT ZN_FING 671 693 C2H2-TYPE.
FT ZN_FING 721 743 C2H2-TYPE.
FT ZN_FING 750 772 C2H2-TYPE.
FT ZN_FING 778 800 C2H2-TYPE.
FT ZN_FING 806 828 C2H2-TYPE.
FT ZN_FING 834 856 C2H2-TYPE.
FT ZN_FING 862 884 C2H2-TYPE.
FT ZN_FING 890 912 C2H2-TYPE.
FT ZN_FING 918 940 C2H2-TYPE.
FT ZN_FING 948 970 C2H2-TYPE.
FT ZN_FING 1010 1038 C2H2-TYPE.
FT ZN_FING 1044 1066 C2H2-TYPE.
FT ZN_FING 1136 1158 C2H2-TYPE.
FT ZN_FING 1164 1186 C2H2-TYPE.
FT ZN_FING 1192 1214 C2H2-TYPE.
FT ZN_FING 1220 1242 C2H2-TYPE.
FT ZN_FING 1248 1270 C2H2-TYPE.
FT ZN_FING 1276 1298 C2H2-TYPE.
FT STRAND 1045 1045

FT STRAND 1052 1052
FT HELIX 1056 1062
FT HELIX 1063 1065
FT TURN 1066 1066
SQ SEQUENCE 1350 AA; 155804 MW; 27F10AB0851E0AD8 CRC64;
Query Match 16.8%; Score 375.5; DB 1; Length 1350;
Best Local Similarity 27.3%; Pred. No. 6,8e-15;
Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;
QY 11 SLTSSSSSRPKYCTCYEGCDKAVNRPFLBQHLRTNSNDREYKCTVDDCAFRKS 70
DB 340 SFLNHQQTHERREKPYLCSH--CNKGFIONSDLVGHFPHRTGERPYQCA--ECHGFIQKS 395
QY 71 HLEHTVSHSEKREPHSCVGVKSVNSQHLKRHITHT--KSPKCTEENCQEAIFYKQSL 128
DB 396 DLVGHLEHTHGEKRFKSHCDKFKTERSAIAKHQRIHTHGEKPYKCS--DQGEFTQRSNL 453
QY 129 -RHHILSVHEKTLCTKQCNKVFTRPSPKLAQHXKH-----HGSPAYQCHRGQPKN 179
DB 454 ILHQRHTGERPRPKCTLCDRFTIONSDLVGHQKVHNLPLSDPHYANSPHKCSK--CDLT 511
QY 180 FQTVSVLQFHFK--QSHPKLCKPCGKCGVCKKGLSSMLSH-----DD 221
DB 512 FSHMSTFEMKSKLHSGEKKFKQCAEKKGFQKSDLVGHILVHIGEKPKLCKKSPSON 571
QY 222 STWKIKW-----TCDYCDVKGFAKKNELVEHNIFPDGNIPLDLKETEVKLENT 272
DB 572 SDLHKHMRHTHGEKPPKPYCTCD--KSFTERSALIKHRT--HTGERPH-----KSCVQCKG 623
QY 273 LDGSKLNNLHLETEKLVKEEDEDEDESDLDKRSQDVRSMSAQR-----S 320
DB 624 FLOKSL-----THSRHTHGEKPYCTQCKSFIONSDLVKHQHLHTHGEKPYHCTEC 676
QY 321 IKSFYASLEGSKSVSKLINSGRK--INCPNNCDRFRSREYDLRRHLKMD--DNLQRTES 378
DB 677 NKRFT--EGSSLVKRRRTTHSGEKPYRCQ--CEKFTFQSSDLVKHLVANGNPNPATA 731
QY 379 FLN-STKEKETPPGEP 393
DB 732 FHEITLRRLNLRSEP 747
RESULT 15
ID 2257 HUMAN STANDARD; PRT; 535 AA.
AC 09Y201;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).
GN ZNF257 OR BMZF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=20054457; PubMed=10585455;
RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
RT hematopoietic cells and identification of a novel transcriptional
RT domain KRNB.";
RL J. Biol. Chem. 274:35741-35748(1999).
CC -1- FUNCTION: May function as a transcription factor.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 1 KRAB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF070651; AAD20957.1; -.
DR HSSP; P08047; ISP2.
DR Genew; HGNC:13498; ZNF257.
DR MIM; 606957; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003677; F:DNA binding; NAS.
DR GO; GO:0008270; F:zinc ion binding; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; Zf-C2H2; 10.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 9.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
KM DOMAIN 4
FT ZN_FING 173 195 KRAB.
FT ZN_FING 201 223 C2H2-TYPE.
FT ZN_FING 229 251 C2H2-TYPE.
FT ZN_FING 257 282 C2H2-TYPE.
FT ZN_FING 288 310 C2H2-TYPE (DEGENERATE).
FT ZN_FING 316 338 C2H2-TYPE.
FT ZN_FING 344 366 C2H2-TYPE.
FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).
FT ZN_FING 403 425 C2H2-TYPE.
FT ZN_FING 431 453 C2H2-TYPE.
FT ZN_FING 459 481 C2H2-TYPE.
FT ZN_FING 487 509 C2H2-TYPE.
SQ SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;

Query Match 16.8%; Score 374; DB 1; Length 535;
Best Local Similarity 33.5%; Pred. No. 3,1e-15;
Matches 88; Conservative 38; Mismatches 115; Indels 22; Gaps 9;

QY 8 KSISLSISSSSRPK-----KYICTYEGCDKAVNRPSSLLEOHLRTHSNDRPYKCTVDDC 62
Db 149 KYVKEFYFESNSDRKIKHTREKTKCKECCGKAFNQSALTTRHKMHTGKEPKYC--EEC 206
QY 63 DKAPFKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENGQCAF 122
Db 207 GKAFNRSSHLTQHKYIHTREPKYCEBCKAFNRSRSHITQHKRIHNRKPPKYDECCKAF 266
QY 123 YKHQSL-----RHHLISVHEKTLTKQCKNVFTRPSKLAQHLKXHHGSSPAYOCDHGCFK 178
Db 267 KWSASALTTLTQHKRIHTGKEPKYCEBCKAFNQSALTTRHKMHTGKEKP-FQCEE--CGK 323
QY 179 NFQTSVLOFHFKQSHPK--LQCPKCGKGVGKGLSSHMLSHDSTMIKIWTCD-YCD 234
Db 324 AFNRSSHLTQHKYIHTREKPKYCEBCKAFNRSRSHITKHKRIH--TREKAYKXDEYCK 379
QY 235 VGKFAKKNELVEHYNIFHDGNIP 257
Db 380 AFNRSSALTTLTQHKIHTGKEKP 402

Search completed: July 22, 2004, 16:46:31
Job time : 14 secs

Query	Best Local Similarity	32.8%	Pred No. 2e-19;
Matches	117;	Conservative	54; Mismatches 129; Indels 57; Gaps 19;
QY	23	KKYITCYEGGCDKAYNPSSLLEQHLRTHNSNDPRPYKCTVDDCDYAFPKSHLEFTHIVSSEK	82
Db	323	KPYKCC--KECGKAFPTOSSSLKTLNLTGKKPYKC--EECDKSPTEKSTLTTLTKRKHITGK	378
QY	83	KPPLHCYVCGGVANSRQHLKRNHELTTH--KSPKCTFENCCEAYVKKQNSLR--HIIISVHEKT	139
Db	379	KPYKCNICGGSPFMSCTNLKTHQTLHTLHGEKPYKC--KECGAALPYMSSLKTHQNLHPGEKP	436
QY	140	LTCQKQKAVTPRPSKLAQHLKLNHNGSGPAPVOCNHPCCFKNFOTWSYLOPH--IKOSHPRKL	197
Db	437	YKCEKCDKSTLLKSLRTHQIHTHGEKP--YKCNIC--CGKSPNCTNLKTHQNLHTHGEKPY	493
QY	198	KCPKCGKGVCGKKGLSSHMLSHDDSTWIKITWCDYCDVGVKAKNLEVHNIVIFPDGNIPT	257
Db	494	KCKCKGKSPHYMSSLKHQNLHSGE--KPYKCEKD--KSPTEKSTLTILKHORI--HTG---	545
QY	258	DDLKLEFTEVKKLE--NLDDGSKLNNHLEFTEKLVVEDEDEDEEDSLDEKSDVYSDSM	315
Db	546	-----KKLYKCNICDK--SFTWCASLKTHK--KPTHGEK-----FY 577	
QY	316	SAORSIKSF--TASLEGSKSVSKLISNSGKKI--NCPANNCNMFSEYDLRRHLKMH	369
Db	578	KCRKCGKSPFOLSTLKHQKIHEDKHTGKTKLYKC--NCCDRSYSHSSRRRKHQLH	632

RESULT 2

08N703 PRELIMINARY; PRT; 404 AA.

ID 08N703

AC 08N703;

DT 01-OCT-2002 (TREMBLrel. 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Hypothetical protein FLJ40479.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eulacerta; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

NP RP

SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shionata N., Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y., Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S., Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R., Kuga N., Kuroda A., Satoh T., Kanata K., Takami S., Terashima Y., Matanabe M., Sugiyama T., Irie R., Otsuki T., Sato K., Nishikawa T., Ishii S., Yamamoto U., Isono Y., Kawai-Hio Y., Salto K., Kimura K., Kimura K., Kanda K., Magatsuna M., Murakawa K., Kanohori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isegai T., "NEBO human cDNA sequencing project"; RT Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL: AK097798; BAC05174.1; -

DR GO: GO:0005634; C:nucleus; IEA.

DR InterPro: IPR007087; Znf_C2H2.

DR InterPro: IPR007086; Znf_C2H2_sub.

DR Pfam: PF00096; zf-C2H2; 14.

DR PRINTS: PR00048; ZINC_FINGER.

DR SMART: SM00355; Znf_C2H2; 14.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.

DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 14.

KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc; zinc-finger.

QY SEQUENCE 404 AA; 45923 MW; B1B35F62F5DF0007 CRC64;

Query Match 17.9%; Score 400; DB 4; Length 404;

Best Local Similarity 33.9%; Pred. No. 6.3e-19;

Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;

QY	23	KKYLCTYEGDCAKAYNP	SLLEQHLRHSNDREYKCTVDDCQAFPRKSHLZTHIVLSSEK	82	
Db	39	KPKKC--EEGKA	FNSSILITTKHKLITHTGEKPYKC--EECGKGFSSVSTLNTKHAITHAE	94	
QY	83	KPHFGSVCGKGVNSR	QHLKRHELTET--KSFCTENCOEAYYKQSLRHILTSVH--EK	138	
Db	95	KPKYCECGKASNSSSK	LMEHKRIHTGEKPYKC--EEGKA	FNSSVSLTEH-KRIHAGEK	151
QY	139	TLVTCOKNKKFTPTPSK	LAQHKLKHGSGSPAYOCDEHGPCFKNQFQWSVLQFH--IKQSHPK	196	
Db	152	PKYCEBGKAFN	SSILITTKHKLITHTGEKPYKC--GCGKA	FNSSVSTLNTKHAITHAE	208
QY	197	LKCPCKCGKGVCGK	KGLSSHMLSHDSTMTIKWTCDYGVGK-FAKNKLVEHYNI	FDHGN	255
Db	209	YKCEBGKASNSSSK	LMEHKRIH--TGKPYKCEEC--GKA	FNSSVSLTEH-KRI-HAE	262
QY	256	IPDDLKETVYKLE	NLLDOGSKYLNLLHLEETKLVAE--DEDEEDSDLEKRDVRS	SDS	314
Db	263	KP--YKCEBGKAF	FTWSSSFTHKKRIHAE-KPYKCEBGKGFSTFSTLITTKHKLITHTGE	318	
QY	315	--MSAQSISFTYAS	LEGSKSVSKLISNSGKKNCPKNNCDMPSPREYDLRRLLKWH	369	
Db	319	KRYKCECGKAF--	SWSSILITTKHKLITHTGEKPYKC--EEGKA	FNSSVSTLNTKHAITH	371
RESULT 3					
Q43693		PRELIMINARY;	PRT;	395 AA.	
ID	Q43693				
AC	Q43693;				
DT	01-JUN-1998	(TReMBLrel. 06, Created)			
DT	01-JUN-1998	(TReMBLrel. 06, last sequence update)			
DT	01-OCT-2003	(TReMBLrel. 25, last annotation update)			
DE	Zinc-finger protein (Fragment).				
GN	ZFRS-1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RX	TISSUE=Testicular tumor;				
RX	MEDLINE=98069676; PubMed=9406578;				
RA	Ogawa T., Poncelet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,				
RA	Udagawa K., Lecocq P., Marine J., Martial J., Hosaka M.;				
RT	"Enhanced expression in seminoma of human zinc finger genes located on				
RT	chromosome 19.";				
RL	Cancer Genet. Cytogenet. 100:36-42(1998).				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).				
DR	EMBL; D70631; BAA24050.1; -.				
DR	HSSP; P08048; 7ZNF.				
DR	GO; GO:0005634; C:nucleus; IEA.				
DR	GO; GO:0008237; F:metallopeptidase activity; IEA.				
DR	GO; GO:0008270; F:zinc ion binding; IEA.				
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR006025; Pept_M_Zn_BS.				
DR	InterPro; IPR007087; ZnF_C2H2.				
DR	InterPro; IPR007086; ZnF_C2H2_sub.				
DR	Pfam; PF00096; ZF-C2H2; 13.				
DR	PRINTS; PR00048; ZINC_FINGER.				
DR	ProDom; PD000003; ZnF_C2H2; 10.				
DR	SMART; SM00355; ZnF_C2H2; 13.				
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.				
DR	PROSITE; PS00157; ZINC_FINGER_C2H2_2; 13.				
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.				
KW	Metal-binding; Nuclear protein; Zinc; Zinc-finger.				
FT	NON TER				
FT	1				
SEQUENCE	395 AA;	45291 MW;	DAE0B54DCE0BF9B CRC64;		
Query Match					
Best local Similarity		17.8%;	Score 397.5;	DB 4; Length 395;	
Matches		97;	Conservative	35; Mismatches 51; Gaps 14;	

QY	23	KYIYCTYECGDYANRPSLLEOHLRHSNDRPYKCTVDDDKAFRKSUHEIIVSHSEK	82
Db	133	KPYKC--KECGAFNNSSLTLTRKRIHTGEXPKYC--EECGKALQSSHLTRKRIHTGE	188
QY	83	KPFHSCVCGKGNVSRQHLKRHEITHT--KSEKTFENCCGAF-----	122
Db	189	KPYKCKCKGKAPNQSMLTTRHEIYHTGEXPKYC--EKGCGAFNHFSLTRKRIHTGEX	246
QY	123	-----YKIQS--LRHHLSVHEKTLTKQCNVFTFRPSKLAOHLKHHGSPAYQC	171
Db	247	YKCKEKGAFKSSLTLTRKRIHTGEXPKYCKEKGAFNQSXLTRHK--KIHGEXPKYEC	305
QY	172	DHGEFGKNQWTASVLOFHIKOSHPK---LKPCPGCGCAGYKKGSLSHMSLHSDSTWIKIM	228
Db	306	EE--CGKAPNQSNLTRH--KXSHTEKPKYKCEGQALNRPSTLTTRKRIH---TGEKPY	359
QY	229	TDYDCDVGK--FAKKNELVEHYNIFHDNTP	257
Db	360	KCEEC--GKAFNQSRLTGKKI--HUGEKP	386

ID	Q81YNO	PRELIMINARY;	PRT;	542 AA.
AC	Q81YNO;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Similar to zinc finger protein 208.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxId=9606;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	T1SSUB=Testis;			
RA	Strausberg R.;			
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, BC035579; AAH35579.1; -			
DR	GO; GO:0005622; C:intracellular; IEA.			
DR	GO; GO:0006376; F:nucleic acid binding; IEA.			
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.			
DR	InterPro: IPR001909; KRAB			
DR	InterPro: IPR007087; Znf_C2H2.			
DR	InterPro: IPR007086; Znf_C2H2_sub.			
DR	Pfam; PF01352; KRAB; 1.			
DR	Pfam; PF00096; zf_C2H2; 11.			
DR	PRINTS; PR00048; ZINC_FINGER..			
DR	ProDom; PD000003; Znf_C2H2; 1.			
DR	SMART; SMO0349; KRAB; 1.			
DR	SMART; SMO0355; Znf_C2H2; 11.			
DR	PROSITE; PS00805; KRAB; 1.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.			
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.			
SQ	SEQUENCE 542 AA; 62756 MW; 335709FE7E094C9B CRC64;			
Query Match	17.6%; Score 392.5; DB 4; Length 542;			
Best Local Similarity	30.0%; Pred. No. 2.7e-18;			
Matches 115; Conservative 39; Mismatches 132; Indels 97; Gaps 19				
QY	8 KSISLSLISSSSSR---PKKICTYEGCDYAVNRPSTLEQLRTHSNDREPKTVDQD 63			
Db	212 KSFCEMLHLTHLQHKRFHTIENSYQC--KDCGKAFFWFSFTLTHRRILHTEKEPKYC--EECG 267			
QY	64 KAFPRKSLHETHIVSHSEKKPEHCSVCGKSGVNSQHLKRHEITLT--KSFKCTFENCQEA 121			
Db	268 KAFNRSSHLLTHKILHTEGKPYKCECGKAFNRSSHLLTHNRILHGVAVPKCT--ECGKA 325			
QY	122 FYKHQSL--RHILLSVEHKTITLCKQCNVFTFPPSKLAQHLKGHGSGPAYQCDHGCENF 180			
Db	326 FNRSSHLLTHRIILHTEGKPYKCECGKAFNQSFTLTHKILHAGEKP--YKCE--CGKAF 382			
QY	181 QTVSVLQFHIKQSHPK--LKCPKCGKGVAGKGLASSHMTSHDSTIMIKITCDYCDVCG 237			

```

Db      363 YRFSLTGH-KTSHTGEKPYKCECGKPNMSSALTQKRIH--TGEPRYKCEB--GK 436
QY      238 -FAKNELVEHYNIPIHDNIPDDLLKETREVKLEJVLDDGSKTANLHLETEKLYAEDE 296
Db      437 AFNESSNLTTH-KMHTGEKP--YKCECGKAFN----- 467
QY      297 EDEEDSLDEKSDVRSDSMSAQRISIKSPYASLEGSKSVSKLISNSGKINCPRNNDWF 356
Db      468 -----RSSQLTAH-----KMHTGEKPYK--EECGKAF 494
QY      357 SREYDLRRH-----LKMHD 370
Db      495 NRSSTLTGHKITHGEKSYKWE 517

```

ID	Q66N38	PRELIMINARY;	PRT;	555 AA.
AC	Q66N38;			
DT	01-DEC-2001	(TREMBLrel. 19, Created)		
DT	01-DEC-2001	(TREMBLrel. 19, last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, last annotation update)		
DS	Hypothetical protein FLJ31444.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,			
RA	Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,			
RA	Okutsu T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,			
RA	Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,			
RA	Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,			
RA	Murakawa K., Kanehori K., Takahashi-Fuji A., Oshima A., Sugiyama A.,			
RA	Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,			
RA	Isoqal T.;			
RT	"MEDO human DNA sequencing project.";			
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
DR	EMBL: AK056006; BAB71072.1; --			
DR	GO: GO:0005654; C:nucleus; IEA.			
DR	GO: GO:0003676; P:regulatory acid binding; IEA.			
DR	GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.			
DR	InterPro: IPR007087; Znf_C2H2.			
DR	InterPro: IPR007086; Znf_C2H2_sub.			
DR	Pfam: PF00096; zfc_C2H2_13.			
DR	PRINTS: PR00048; ZINC_FINGER.			
DR	ProDom: PD000003; Znf_C2H2_1.			
DR	SMART: SM00349; KRAB; 1.			
DR	SMART: SM00355; Znf_C2H2_13.			
DR	PROSITE: PS50805; KRAB; 1.			
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.			
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 14.			
KW	Hypothetical protein; Metal-binding; Nuclear protein; zinc;			
KW	zinc-finger.			
SQ	SEQUENCE 555 AA; 64025 MW; 75C246D4820FB920 CRC64;			
Query Match	17.6%;	Score 392;	DB 4;	Length 555;
Best Local Similarity	32.5%;	Pred. No. 3e-18;		
Matches 116;	Conservative 45;	Mismatches 162;	Indels 34;	Gaps 17;
QY	21 RPKKYLCTYEGCDKAVNRPSTLEQRIHRTNSDRPYKTVDDCDVAFPKSHLETHIVSHS 80			
Db	169 RENSYYG-C-EECGKVFRRSTLRLRHKKVHNGEAPFC-EECGKAFHSSLTLLTHKKIHT 224			
OY	81 EKRPFGCSVCGKQVNSRQHLKRIHTYT-KSKRCTFENCOEAPYKHQST-RHHLSVHE 137			
Db	225 GEXPKVCECGKAFYHSSHLLTHKVIHTGKPKFC-EECGKAFNHPASLTLTKFTIHVE 282			
Y	138 KTLTLCQCKNVFRRPSKLAQHLKHHGGSAPAYCCDHDGCEKRNQTSNVIQFH-ITQSHD 195			

Db 559 G6KPYCCEGKAFNNSSNLTK_HN1IHGKSYKC--EEGKAFNQSSTLTTRK 611

RESULT 8

ID	Q8NR04	PRELIMINARY;	PRT;	592 AA.
AC	Q8NR04;			
DT	01-OCT-2002 (T-EMBLrel. 22, Created)			
DT	01-OCT-2002 (T-EMBLrel. 22, Last sequence update)			
DE	Hypothetical protein FLJ39023.			
OS	Homo sapiens (human) .			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Nimnuya K., Magasuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,			
RA	Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kanihara K.,			
RA	Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,			
RA	Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto U., Isono Y.,			
RA	Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamaishi H.,			
RA	Matsuo K., Nakamura Y., Sekine M., Kiyuchi H., Murakawa K.,			
RA	Kanohori K., Takahashi-Fujii A., Ohnishi A., Sugiyama A., Kawakami B.,			
RA	Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,			
RT	"NEO" human cDNA sequencing project."			
RL	Submitted (JUN-2002) to the EMBL/Genbank/DBS databases.			
CC	-1 SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).			
CC	EMBL; AK096342; BAC04764.1; -			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0008237; F:metalloproteinase activity; IEA.			
DR	GO; GO:0003676; F:nucleic acid binding; IEA.			
DR	GO; GO:0008270; F:zinc ion binding; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.			
DR	InterPro; IPR001909; KRAB.			
DR	InterPro; IPR006025; Pept M_zn_BS.			
DR	InterPro; IPR007087; Znf_C2H2.			
DR	InterPro; IPR007086; Znf_C2H2_sub.			
DR	Pfam; PF01352; KRAB. 1.			
DR	Pfam; PF00096; zf-C2H2; 15.			
DR	PRINTS; PR00048; ZINC_FINGER.			
DR	ProDom; PD000003; Znf_C2H2; 5.			
DR	SMART; SM00349; KRAB. 1.			
DR	SMART; SM00355; ZNF_C2H2; 16.			
DR	PROSITE; PS00805; KRAB. 1.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.			
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 16.			
DR	PROSITE; PS00142; ZINC_PROTEAS; 4.			
KW	Hypothetical protein; Metal-binding; Nuclear protein; Zinc;			
KW	Zinc-finger.			
SQ	SEQUENCE 592 AA; 67920 MW; D116FB9366EBABDE CRC64;			
Query Match	17.5%; Score 391; DB 4; Length 592;			
Best Local Similarity	31.1%; Pred. No. 3.8e-18;			
Matches 120; Conservative 41; Mismatches 161; Indels 64; Gaps 19				
QY	23 KKYICTYGGCKAVRPSLLEQHRTHNDSEPKYQTVDDCKAFPRKSHLEHIVSHSEK 82			
Db	199 KPYIC--EEGKAFKYSALTHTKRIRHTEBEKYNCKGKAFIASSTLSRRHFIM--GKKHYK 371			
QY	83 KPHKCVGKGVNSRHLKREHITTT--KSEKCTENCOEAFYKQSLRH----- 131			
Db	255 KPYKCEEGKAFNQSSTLTTRKIRHTEBEKYPKC--EEGKAFNQSSTLTTRKIRHTEBEK 312			
QY	132 -----ILSVH-----EKLTLTKQCNKVFTRPSKLAQHLKHHGSPAYOC 171			
Db	313 VVCEEGKAFKYSRLTLTKRIRHTEBEKYNCKGKAFIASSTLSRRHFIM--GKKHYK 371			
QY	172 DHPCEKFNKFTWSTVLOFH--IKQSHPRKLCPRGCGKGGKGLSHMLSHDSTPMIKIT 229			
Db	372 EE--CGKAFIMSSVLTTRKIRHTEBEKYPKC--EEGKAFNQSSTLTTRKIRHTEBEK 426			

QY	230	CDYCVWCK-FAKNELVEHYVNFHDSNIDDDLKFEVKKLENLDQSKLNNLHELE7-287
Db	427	CEEC-GKAFNOSSSLTKHK1-HTGEKE---YKCECGKAFN---QSSSLTKHKK1HTG477
QY	288	-EKLYVEDEEEDS---LDEKSDVSDSDMSAQSISFPAISLEGSYSVSLJNSGK343
Db	478	KFPYKCEBGKAFNOSTLTKHK1HTREKFPYKCECGKAF-HLSTHLTKK1HTGEK535
QY	344	KINCPKNNCDPMFSREYDLRRHLKWA369
Db	536	PYRC--RECCKAFNHSATLSHKKH559
RESULT	9	
QBN211		
ID	QBN211	PRELIMINARY; PRT; 540 AA.
AC	QBN211;	
DT	01-OCT-2002 (TREMBLrel. 22, Created)	
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)	
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)	
DE	Hypothetical protein FlvJ36350.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Metazoa; Primates; Catarrhini; Homidae; Homo.	
OX	NCBI_TaxID=9606;	
RP	[1]	
RC	SEQUENCE FROM N.A.	
RA	TISSUE=Thymus;	
RA	Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,	
RA	Arita M., Muehshino K., Yuuki H., Hara H., Sugiyama T., Irie K.,	
RA	Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,	
RA	Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,	
RA	Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,	
RA	Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.,	
RT	"NEDO human cDNA sequencing project."	
RL	Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.	
CC	-i- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).	
EMBL	AK093669; BAC04216.1; -.	
DR	PIR; F42075; F42075.	
DR	GO; GO:0005634; C:nucleus; IEA.	
DR	GO; GO:0003676; F:nucleic acid binding; IEA.	
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	
DR	InterPro; IPR001909; KRAB.	
DR	InterPro; IPR007087; Znf_C2H2.	
DR	InterPro; IPR007086; Znf_C2H2_sub.	
DR	Pfam; PF01352; KRAB; 1.	
DR	Pfam; PF00096; zf-C2H2; 12.	
DR	PRINTS; PRO0048; ZINC_FINGER.	
DR	ProDom; PD000003; Znf_C2H2; 1.	
DR	SMART; SM00349; KRAB; 1.	
DR	SMART; SM00355; Znf_C2H2; 12.	
DR	PROSITE; PS50805; KRAB; 1.	
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.	
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.	
KW	Hypothetical protein; Metal-binding; Nuclear protein; Zinc;	
KW	zinc-finger.	
SQ	SEQUENCE 540 AA; 62985 MW; 84D72EC15A7002C CRC64;	
Query Match	17.5%;	Score 390; DB 4; Length 540;
Best Local Similarity	30.0%;	Pred. No. 4e-18;
Matches 107;	Conservative 61;	Mismatches 155; Indels 34; Gaps 15;
QY	23	KKYICTEGCDKAVNPSLLEOHILRTSNRPYKCTVDDCDKAFRKSHTLTHVSHSEK82
Db	198	KLYKC-QBEDRITNQSNSLTKYDAREKPYKC--EEGKAFNOSSSLTKHK1HTGE253
QY	83	KPFHGVGVGKAVNSROHLKKEHILTHYSFKCTFENCOEAFYKQSL-RHHILSVHEKTLT141
Db	254	KPYKCECGKAFNOSNLTKHK1HTGEOPYPICEGKAFLOSTLTTHKRIHNGEKPKY313
QY	142	CKQCNKVFTRPSKLAQKHLNHHGSPRAYQCDNHEGCFNPFQTVSLQTH--IKOSHPRKLC199

Db 314 CEECGAFNRSSKLTLEHKNHITGEOP--YKCEE--CGKAFNRSSNLTEHKKITHTEKPYKC 370

Qy 200 PKCGKCGVCKGKGLSSMLSHDSTWTKIWTCDYGVCK--FAKKNELVEHYNITHDGNIPD 258

Db 371 KEECGAFKSHSSALTTKRIH---TGEKPYKCEEC--GKAFNRSSKLTLEHKKI--HTGK---421

Qy 259 DLKETEYKLTENLLDQSGKLNHLTEKLTAVEDE-----EDEEDSLDERKSPVRS 313

Db 422 ---KPYKCECGKAFQSSKLTLEHKNHSGEIPYKCECGKAFKSHSSALTTKRIHTEK 478

Qy 314 SMSAQRISKFTASLEGSKSVSKLISNSGKK-INCPKNCDFRFRPYDLRHLKWM 369

Db 479 PYKCECGKAFNRSS---SKLTLEHKNHITGEKPYKCEC--CDKAFNQSANLTGKKIH 530

RESULT 10

Q96N22 PRELIMINARY; PRT; 576 AA.

AC 096N22;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Hypothetical protein FJ331526.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Niinomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,

RA Niinuma T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,

RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto U., Isono Y.,

RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,

RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,

RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isegai T.,

RT "MDO human cDNA sequencing project.";

RU Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; AK056088; BABY1090.1; -

DR GO; GO:0005634; C:nucleus; IEA.

DR InterPro; IPR007087; Znf_C2H2.

DR InterPro; IPR007086; Znf_C2H2_sub.

DR Pfam; PF00096; Zf-C2H2_15.

DR PRINTS; PR00048; ZINC_FINGER.

DR ProDom; PD000003; Znf_C2H2_2.

DR SMART; SM00355; Znf_C2H2_16.

DR PROSITE; PS00026; ZINC_FINGER_C2H2_1; 10.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 16.

KM Hypothetical protein; Metal-binding; Nuclear protein; zinc;

KW Zinc-finger.

SQ SEQUENCE 576 AA; 66620 MW; 4396672D34BP99CD CRC64;

Query Match 17.5%; Score 390; DB 4; Length 576;

Best Local Similarity 28.6%; Pred. No. 4.3e-18;

Matches 119; Conservative 57; Mismatches 166; Indels 72; Gaps 17;

Qy 3 ESDFTKSISSSSSRPKYITCTYGCDAVNRPELHQLHTSHNDPRKYTVDC 62

Db 217 ECDKAFNQSSTLTTHKILITREKLINEYKECGKAFNQSHTLTHKILITGEKPYKC--EEC 274

Qy 63 DKAFRKSHETHIVSHSEKKPFHSCVCGKGVNSQHLKREHITPT--KSPKCFEENCQ 120

Db 275 GKAFNQSHTLTHKILITGEKPYKCECGKAFNQSHTLTHKILITGEKPYKC--EECGK 332

Qy 121 AFYKHQS--RHILSVHEKTLTKCOCNVFTFPPSLAQHKLKHGSGPAYOCDFGCFKN 179

Db 333 AFNKSHTLTHKILITGEKPYKCECGKAFNQSHTLTHKILITGEKPYKC--EECGK 389

Qy 180 FQTVSLVDFH--IKSHPKLCKPCGKCGVCKGKGLSSMLSHDSTWTKIWTCDYGVCK 237

Db 390 FNFQSNLTTHKRIHTEKPYKCECGKAFNQSSTLTTHKRIH---TGEKSYKCEEC--GK 444

Qy 238 -FAKKNELVEHYNIFPDGNIPDILLKETEYKLTENLLDQSGKLNHLTEKLTAVEDE 296

Db 445 AFYRSSKLTLEHKKI--HTGEK-----YTGECKKAFNHSALTAHIV--IHTGE 490

Qy 297 EDEEDSLDERKSPVRSDSMSAQRISKFTASLEGSKSVSKLISNSGKINCPCNCDRMP 356

Db 491 KPYQ-----CECGKAFNQSHTLR--HKRIHTEKPYKCEC--CGKAF 530

Qy 357 SREYDLRHLKXWDDNLQRIESFLNISEKETEYKLTAVEDEYKAMMDLIPNFTSVYSR 412

Db 531 NQSSNLTHKRIH-----TGEKLYPKKCNDFPNTSKFSK 566

RESULT 11

Q9VG72 PRELIMINARY; PRT; 501 AA.

AC Q9VG72;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE CG5245 protein.

GN CG5245.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OC NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.F., Bhandari D., Bolshakov S.,

RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laake F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclel B.J.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert R., Remington K., Saunders R.D.C., Scheelers F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskis R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yen R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,

RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.N., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL EMBL; AE003696; AA54813.1; -

DR HSSP; P08047; 1SR2.

DR FlyBase; FBgn0038047; CG5245.


```

Query Match      17.4%; Score 387; DB 11; Length 559;
Best Local Similarity 28.7%; Pred. No. 6.6e-18;
Matches 109; Conservative 55; Mismatches 136; Indels 80; Gaps 17;

QY 23 KKYICTYEGCDKAYNRPSLLLEQHLRTHSNDPRPKCTVDDCDKAFPRKSHLETHIVSHSEK 82
DB 187 KPYKCNQ--CDKAFSGHSTLTQTHRRTHHTGKPKYC--DQCDKAFSEKCSLTQTHRRTHHTG 242
QY 83 KPFHSCVCGKGVNSRQHLKRHEITHT--KSFKC-----TFEN----- 117
DB 243 KPYKCNQCDKAFSGYSHLTHRRTHHTGKPKYKCNCEDETFNSNSNLTQTHRRTHHTGKPKYK 302
QY 118 ---COAFYKHOSLRHHILTS--VHEKTLTKCKCNKVFTRPSKLAQHLKHHGSPAYQCDH 173
DB 303 CNQCDKAFSGHSTLTQTHRRTHHTGKPKYKCNQCDKAFSRHSTLTQTHRRTHHTGKPKYK 361
QY 174 PGCFKNPOTWSVLQFIHKQSH---PKLCPKCGKGVCGKKGLSSHMLSHDSTMIKIWTG 230
DB 362 --CDKAFSGYSHLTH--RTHHTGKPKYKCNQCNKAFSGYSHLTHRRTH---TGEKPKYC 415
QY 231 DYCDVGFPAKKNELVEHYNIFFHGNITPDDILKETEVKLENLDDQSKLNNLHELETKL 230
DB 416 NQCD--KTFPSNHSHTLTQTHRRTHHTGK-----KPYKCNQCDKAFSRHSTLTQTHRRTHHTGK 467
QY 291 KVEDEDEDESDLSDEKSDVSDMSAQRISKFTASLSEGSYSKLSINSKGI--NCPK 349
DB 468 PFKCNQCDK-----AFSQKCSLTQKHRIHT-----GKTLKNC-- 499
QY 350 NNCDRMFSREYDLRRLKWH 369
DB 500 NECCKAFSGHSTLTQTHRRTH 519

RESULT 14
Q8BJ46 PRELIMINARY; PRT; 297 AA.
ID O8BJ46
AC O8BJ46;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE MSZF33 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=1246851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK032220; BAC27766.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; Zf-C2H2; 8.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
SQ SEQUENCE 297 AA; 34666 MW; 2B94CC27CB2635EB CRC64;

Query Match      17.3%; Score 385.5; DB 11; Length 297;
Best Local Similarity 38.1%; Pred. No. 4.2e-18;
Matches 93; Conservative 36; Mismatches 82; Indels 33; Gaps 14;

QY 23 KKYICTYEGCDKAYNRPSLLLEQHLRTHSNDPRPKCTVDDCDKAFPRKSHLETHIVSHSEK 82
DB 69 KRYKCNQ--CDKAFSRHSTLTQTHRRTHHTGKPKYC--NQCGAFYQSHLTHHTVTHHTGE 124
QY 83 KPFHSCVCGKGVNSRQHLKRHEITHT--KSFKCTFENCOEAFYKHOSLRHHILTSVHEKTL 140

```

```

DB 125 KPYKCNQCDKAFACHNKLQKHRRTHHTGKPKYC--DQCKAF-----VYESYLQVHKSH 177
QY 141 T-----CNQCNKVFTRPSKLAQHLKHHGSPAYQCDHGPCKNPQTSVLQFIHKOSH 194
DB 178 TGEKPKYKCNCGGAFRSHLKHVKHTTHTEKPKYKCNQ--CGKALAHYSTLQVH--QRTH 233
QY 195 PKL--CPKCGKGVCGKKGLSSHMLSHDSTMIKIWTCDYCDVGF--FAKKNELVEHYNI 250
DB 234 TGEKPYECQCGGAFANQGYFYQHKRIH---TGEKPKYKCDQ--GKAFVSSDLKRHERV 288
QY 251 FHDG 254
DB 289 -HTG 291

RESULT 15
Q8VECI PRELIMINARY; PRT; 511 AA.
ID Q8VECI
AC Q8VECI;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Similar to zinc finger protein 40.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 11
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC019219; AAH19219.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 12.
DR SMART; SM00355; Znf_C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 511 AA; 59582 MW; 11C75B538FZDB6F CRC64;

Query Match      17.1%; Score 382; DB 11; Length 511;
Best Local Similarity 31.1%; Pred. No. 1.3e-17;
Matches 110; Conservative 47; Mismatches 143; Indels 54; Gaps 17;

QY 18 SSSRPKKYICTYEGCDKAYNRPSLLLEQHLRTHSNDPRPKCTVDDCDKAFPRKSHLETHIV 77
DB 138 TNNREKTYKCS--EDCKCFYKCKLRQHRHTGKPKYKCS--EDCKCFYKCKLRILIHQR 193
QY 78 SHSEKPFHSCVCGKGVNSRQHLKRHEITHT--KSFKCTFENCOEAFYKHOSLRHHILTSV 135
DB 194 VHTGEKPKYKCSBGDKGFTSKSRILNIHQRVHTGKPKYKCS--EDCKSFQOQNLSTIH-LRI 250
QY 136 H--EKTITCKQCNKVFTRPSLAQHLKHHGSPAYQCDHGFENPQTSVLQFIH--IX 191
DB 251 HTGKPKYKCSBGDKCFYKSGLRSHQI--HTGEKAFYKCSF--CGKCFYHKQDLRKHQSIH 307
QY 192 QSHPKLCPKCGKGVCGKKGLSSHMLSHDSTMIKIWTCDYCDVGFPAKKNELVEHYNIFF 251
DB 308 MGEKPKYKCSBGDKGFTQKRLIITHQIHTGKPKYKCSBGDKC-----FLKSDLSIHQI- 362
QY 252 HDGNIPDDLKETEVKLENLDDQSKLNNLHELETKLVEBDEDESDLSDEKRSYVR 311
DB 363 HTGE-----KPYECSECGGAFANQGYFYQHKRIH---TGEKPKYKCDQ--GKAFVSSDLKRHERV 395
QY 312 SDSMSAQRISKFTASLSEGSYSKLSINSKGI--NCPKNDPRMFSREYDLRRLH 365
DB 396 --SYKSECDKSFQ--QGNLRILHRIHTGKPKYC--SECKCFYKSGLRSH 443

Search completed: July 22, 2004, 16:47:24
Job time : 43 secs

```